

PARTS

Die Evaluation eines multimodalen Präventionsprogramms zur Förderung von Intergruppenbeziehungen bei Grundschulkindern

Dissertation

zur Erlangung des akademischen Grades

Doctor Philosophiae (Dr. phil.)

vorgelegt dem Rat der Fakultät für Sozial- und Verhaltenswissenschaften

der Friedrich-Schiller-Universität Jena

von Dipl.-Psych. Michael Saur

geboren am 17.10.1977 in Datteln

Gutachter

1. Prof. Dr. Andreas Beelmann

2. Prof. Dr. Peter Noack

Datum der mündlichen Prüfung: 04.07.2011

Für meine Eltern

Vorwort

Die vorliegende Arbeit wurde im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projektes „Förderung von Intergruppenbeziehungen bei Kindern“ (BE 3731/1-1 und BE 3731/1-2) der DFG-Forschergruppe „Discrimination and Tolerance in Intergroup Relations“ (FOR 481) erstellt. Die Projektlaufzeit betrug fünf Jahre, von Oktober 2005 bis November 2010. In diesem Zeitraum wurde das Interventionsprogramm PARTS – Programm zur Förderung von Akzeptanz, Respekt, Toleranz und Sozialer Kompetenz – entwickelt, durchgeführt und evaluiert. Die Evaluation von PARTS ist das Thema dieser Arbeit.

Viele Personen waren am Erfolg des Projekts und an der Entstehung der Daten für diese Arbeit beteiligt. An dieser Stelle möchte ich mich bei den beteiligten Personen ganz herzlich für ihre Mitarbeit und Unterstützung bedanken!

Allen voran bin ich natürlich dem Leiter und Initiator des Projekts Prof. Dr. Andreas Beelmann zu Dank verpflichtet! Auch bei Prof. Dr. Peter Noack bedanke ich mich für die gute Zusammenarbeit in unserem Kooperationsprojekt. Außerdem dürfen die Projektmitarbeiterinnen Diana Schulze und Petra Ziegler in meiner Danksagung natürlich nicht fehlen.

Ganz besonders möchte ich die Arbeit und Hilfe zahlreicher Hilfskräfte betonen, ohne deren Einsatz und Motivation dieses Projekt nie so erfolgreich hätte verlaufen können. Hier sind besonders hervorzuheben: Ulrike Bogsch, Rebekka Carl, Julia Küchel, Antje Becker und Anna Steudel.

Und natürlich möchte ich allen Kindern, Familien, Lehrern und Schulleitern danken, die sich so kooperativ, hilfsbereit und begeistert an diesem Projekt beteiligt haben.

Inhaltsverzeichnis

I.	Einleitung	15
II.	Theorie	19
II.1.	Grundlagen	19
II.1.1.	Gruppenbezogene soziale Einstellungen	19
II.1.2.	Toleranz	23
II.1.3.	Ausdrucksformen gruppenbezogener sozialer Einstellungen	25
II.2.	Die Entwicklung gruppenbezogener sozialer Einstellungen	27
II.2.1.	Der Entwicklungsverlauf von Vorurteilen und Stereotypen	27
II.2.2.	Entwicklungstheorien	29
II.3.	Empirische Studien	47
II.3.1.	Klassifikation von Interventionsansätzen	47
II.3.2.	Kontakt-Interventionen	51
II.3.3.	Stellvertretender-Kontakt-Interventionen	55
II.3.4.	Wissensbasierte Interventionen	68
II.3.5.	Förderung individueller Kompetenzen	80
II.4.	Schlussfolgerungen für PARTS	95
II.4.1.	Schlussfolgerungen für die Planung und Durchführung von PARTS ..	95
II.4.2.	Schlussfolgerungen für Zielkriterien von PARTS	99
III.	Methode n	102
III.1.	Konzeptionelle Grundlagen von PARTS	102
III.2.	Ablauf und Durchführung der Intervention	104
III.2.1.	Gesamtablauf	104
III.2.2.	Rahmenbedingungen der Programmdurchführung	107
III.2.3.	Umsetzung der Interventionsstrategien im Rahmen von PARTS	109
III.3.	Design der Untersuchung	120

III.4.	Die Stichprobe.....	122
III.4.1.	Rekrutierung der Stichprobe	122
III.4.2.	Entwicklung der Stichprobe.....	124
III.4.3.	Stichprobenpflege	125
III.4.4.	Fallausschlüsse für die Evaluationsstichprobe	126
III.4.5.	Beschreibung der Evaluationsstichprobe	127
III.4.6.	Beschreibung der Datenlage.....	133
III.5.	Durchführung der Untersuchungen	134
III.5.1.	Der Prätest T1	134
III.5.2.	Der Posttest T2	136
III.5.3.	Das Follow-up T3	137
III.6.	Messinstrumente	141
III.6.1.	Implementationsmaße.....	141
III.6.2.	Kognitive und sozial-kognitive Konstrukte	142
III.6.3.	Gruppenbezogene soziale Einstellungen	150
III.6.4.	Gruppenbezogene soziale Einstellungen	
	mit kognitivem Schwerpunkt.....	151
III.6.5.	Gruppenbezogene soziale Einstellungen	
	mit affektivem Schwerpunkt	155
III.6.6.	Gruppenbezogene soziale Einstellungen	
	mit konativem Schwerpunkt	158
III.6.7.	Interpersonale soziale Einstellungen und Sozialverhalten	159
III.7.	Operationalisierung der Ziele von PARTS.....	165
III.8.	Hypothesen der PARTS-Studie.....	169
IV.	Ergebnisse	171
IV.1.	Allgemeine Auswertungsstrategien	171
IV.2.	Statistische Methoden	174
IV.2.1.	Deskriptive Statistik	174
IV.2.2.	Inferenzstatistik	175
IV.2.3.	Voraussetzungen	
	der angewendeten inferenzstatistischen Verfahren	176

IV.2.4.	Prüfung der Voraussetzungen und	
	Umgang mit Voraussetzungsverletzungen.....	177
IV.3.	Implementationsanalyse.....	180
IV.3.1.	Implementation des Gesamtprogramms	180
IV.3.2.	Implementation getrennt nach Interventionsbereichen.....	186
IV.3.3.	Gruppe der „Störenfriede“	188
IV.3.4.	Implementation unter Einbezug der Kovariaten	191
IV.3.5.	Implementationsanalyse – Zusammenfassung	196
IV.4.	Wirksamkeitsanalyse.....	198
IV.4.1.	Ergebnisse zu den proximalen Effektmaßen.....	198
IV.4.2.	Ergebnisse zu den distalen Effektmaßen.....	204
IV.4.3.	Gruppenbezogene soziale Einstellungen	
	mit kognitivem Schwerpunkt.....	204
IV.4.4.	Gruppenbezogene soziale Einstellungen	
	mit affektivem Schwerpunkt	215
IV.4.5.	Gruppenbezogene soziale Einstellungen	
	mit konativem Schwerpunkt	224
IV.4.6.	Interpersonale soziale Einstellungen.....	227
IV.4.7.	Wirksamkeitsanalyse – Zusammenfassung	235
V.	Diskussion.....	238
V.1.	Ziele der vorliegenden Arbeit	238
V.2.	Anforderungen an ein wissenschaftlich fundiertes Programm.....	240
V.3.	Einordnung der Ergebnisse zu strukturellen Merkmalen von PARTS	243
V.3.1.	Inhaltlicher Aufbau des Programms	243
V.3.2.	Intensität und Dauer des Programms.....	245
V.4.	Zusammenfassung und Erörterung der Implementationsanalyse	248
V.4.1.	Implementation allgemein	248
V.4.2.	Implementation einzelner Interventionsbereiche	249
V.4.3.	Identifikation der Störenfriede	251
V.4.4.	Implementation abhängig von Personenmerkmalen	251
V.4.5.	Fazit zur Implementationsanalyse.....	256
V.5.	Zusammenfassung und Erörterung der Wirksamkeitsanalyse	257

V.5.1.	Zusammenfassung und Einordnung der Ergebnisse	
	zu den proximalen Effektmaßen.....	257
V.5.2.	Zusammenfassung und Einordnung der Ergebnisse	
	zu den distalen Effektmaßen.....	267
V.6.	Kernpunkte der Befunde, Erläuterungen und Zielerreichung von PARTS.....	285
V.7.	Grenzen und Möglichkeiten der Arbeit	292
V.8.	Implikationen für Forschung und Praxis	298
VI.	Zusammenfassung	301
VII.	Literatur	303

Tabellenverzeichnis

Tabelle 01: Abgrenzung der Begriffe Toleranz, Akzeptanz und zwei Formen des Respekts	25
Tabelle 02: Klassifikation von Interventionen zur Prävention negativer gruppenbezogener sozialer Einstellungen	49
Tabelle 03: Gesamtablauf von PARTS	105
Tabelle 04: Stellvertretender-Kontakt-Intervention bei PARTS	111
Tabelle 05: Wissensbasierte Interventionen bei PARTS	115
Tabelle 06: Förderung sozial-kognitiver Kompetenzen bei PARTS	118
Tabelle 07: Teilnahme und Ausfallraten an den PARTS-Untersuchungen, bezogen auf die Teilnehmer zu t1	125
Tabelle 08: Fallausschlüsse und Ausfälle zur Evaluationsstichprobe	127
Tabelle 09: Geschlecht und Alter der Evaluationsstichprobe über die Zeit	129
Tabelle 10: Vergleich soziodemographischer Variablen der Mütter nach Interventions- und Kontrollgruppe in Prozent	131
Tabelle 11: Vergleich soziodemographischer Variablen der Väter nach Interventions- und Kontrollgruppe in Prozent	132
Tabelle 12: Datenlage in der Evaluationsstichprobe	133
Tabelle 13: Instrumente der Gruppen- und Einzelbefragung der Kinder nach Messzeitpunkten	139
Tabelle 14: Instrumente der Befragung der Eltern und Lehrer nach Messzeitpunkten	140
Tabelle 15: Interne Konsistenzen der Interne Konsistenzen der Skalen der Gruppen- und Einzelbefragung der Kinder	163
Tabelle 16: Interne Konsistenzen der SDQ-Skalen getrennt nach Informanten zu t1, t2 und t3	164
Tabelle 17: Hypothesen der PARTS-Studie zu proximalen Effektmaßen	169
Tabelle 18: Hypothesen der PARTS-Studie zu distalen Effektmaßen	170
Tabelle 19: On-Task-Verhalten der Teilnehmer über die 15 Lektionen, getrennt nach Lehrer- und Trainereinschätzungen, sowie Korrelationen der Einschätzungen von Lehrern und Trainern	182

Tabelle 20: Off-Task-Verhalten der Teilnehmer über die 15 Lektionen, getrennt nach Lehrer- und Trainereinschätzungen, sowie Korrelationen der Einschätzungen von Lehrern und Trainern	183
Tabelle 21: Spaß der Teilnehmer über die 15 Lektionen, getrennt nach Lehrer- und Trainereinschätzungen, sowie Korrelationen der Einschätzungen von Lehrern und Trainern	184
Tabelle 22: Mittelwerte, Standardabweichungen, Effektstärken und 95%-Konfidenzintervalle, nach Interventionsbereichen getrennt, im Vergleich zum Gesamtmittelwert der Implementationsparameter	185
Tabelle 23: Anzahl der Kinder mit hohem Off-Task-Verhalten, getrennt nach Lehrer und Trainereinschätzungen, über die 15 Lektionen hinweg	190
Tabelle 24: Geschlechtsunterschiede der mittleren Implementationsparameter getrennt nach Lehrer- und Trainereinschätzungen	191
Tabelle 25: Korrelationen zwischen dem Alter der Kinder zum ersten Messzeitpunkt und den mittleren Implementationsparametern	192
Tabelle 26: Mittlere Implementationsparameter für Stadt- versus Landkinder getrennt nach Lehrer- und Trainereinschätzungen	193
Tabelle 27: Korrelationen zwischen der Intelligenz der Kinder zum ersten Messzeitpunkt und den mittleren Implementationsparametern	194
Tabelle 28: Mittlere Implementationsparameter für Störenfriede versus Nicht-Störenfriede, getrennt nach Lehrer- und Trainereinschätzungen	195
Tabelle 29: Mittelwerte, Standardabweichungen und Netto-Effektstärken des interkulturellen Wissens getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	199
Tabelle 30: Mittelwerte, Standardabweichungen und Netto-Effektstärken der multiplen Klassifikationsfähigkeit mit sozialen Stimuli, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	200
Tabelle 31: Mittelwerte, Standardabweichungen und Netto-Effektstärken der Perspektivenübernahmefähigkeit, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	201
Tabelle 32: Mittelwerte, Standardabweichungen und Netto-Effektstärken des Tests zum sozialen Problemlösen, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	202

Tabelle 33: Mittelwerte, Standardabweichungen und Netto-Effektstärken der MRA-Eigengruppen-bevorzugung (MRA-EGB) gegenüber einem russischem Kind (r), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	205
Tabelle 34: Mittelwerte, Standardabweichungen und Netto-Effektstärken der MRA-Fremdgruppen-abwertung (MRA-FGA) gegenüber einem russischen Kind (r), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	206
Tabelle 35: Mittelwerte, Standardabweichungen und Netto-Effektstärken der kognitiven Verkuyten-Skala gegenüber russischen Kindern (r), getrennt nach Interventionsgruppe und Kontrollgruppe zum dritten Messzeitpunkt	207
Tabelle 36: Mittelwerte, Standardabweichungen und Netto-Effektstärken der MRA-Eingengruppenbevorzugung (MRA-EGB) gegenüber einem chinesischen Kind (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	208
Tabelle 37: Mittelwerte, Standardabweichungen und Netto-Effektstärken der MRA-Fremdgruppenabwertung (MRA-FGA) gegenüber einem chinesischen Kind (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	209
Tabelle 38: Mittelwerte, Standardabweichungen und Netto-Effektstärken der kognitiven Verkuyten-Skala gegenüber chinesischen Kindern (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	211
Tabelle 39: Mittelwerte, Standardabweichungen und Netto-Effektstärken des IBM-Sympathie-Items gegenüber einem russischen Kind (r), getrennt nach Interventionsgruppe und Kontrollgruppe zum dritten Messzeitpunkt....	215
Tabelle 40: Mittelwerte, Standardabweichungen und Netto-Effektstärken des Verkuyten-Sympathie-Items gegenüber der Gruppe russischer Kindern (r), getrennt nach Interventionsgruppe und Kontrollgruppe zum dritten Messzeitpunkt	216
Tabelle 41: Mittelwerte, Standardabweichungen und Effektstärken der Haus-Aufgabe gegenüber einer russischen (r) Familie, getrennt nach	

Interventionsgruppe und Kontrollgruppe zu den Messzeitpunkten t2 und t3	217
Tabelle 42: Mittelwerte, Standardabweichungen und Netto-Effektstärken des IBM-Sympathie-Items gegenüber einem chinesischen Kind (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	218
Tabelle 43: Mittelwerte, Standardabweichungen und Netto-Effektstärken des Verkyuten-Sympathie-Items gegenüber der Gruppe chinesischer Kinder (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	220
Tabelle 44: Mittelwerte, Standardabweichungen und Effektstärken der Haus-Aufgabe gegenüber einer chinesischen (c) Familie, getrennt nach Interventionsgruppe und Kontrollgruppe zu den Messzeitpunkten t2 und t3	221
Tabelle 45: Mittelwerte, Standardabweichungen und Netto-Effektstärken der IBM-Kontaktintention-Eigengruppenbevorzugung (IBM-K-EGB) gegenüber einem russischen Kind (r), getrennt nach Interventionsgruppe und Kontrollgruppe zum dritten Messzeitpunkt	224
Tabelle 46: Mittelwerte, Standardabweichungen und Netto-Effektstärken der IBM-Kontaktintention-Eigengruppenbevorzugung (IBM-K_EGB) gegenüber einem chinesischen Kind (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	225
Tabelle 47: Mittelwerte, Standardabweichungen und Netto-Effektstärken des Fragebogens zur interpersonalen Toleranz, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	228
Tabelle 48: Mittelwerte, Standardabweichungen und Netto-Effektstärken des SDQ, eingeschätzt durch die Eltern, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	230
Tabelle 49: Mittelwerte, Standardabweichungen und Netto-Effektstärken des SDQ, eingeschätzt durch die Lehrer, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg	232
Tabelle 50: Zusammenfassung kurzfristiger und langfristiger Interventionseffekte der proximalen Maße	235

Tabelle 51: Zusammenfassung kurzfristiger und langfristiger Interventionseffekte der distalen Maße.....	236
---	------------

Abbildungsverzeichnis

Abbildung 01:	Modell der Integrative Developmental-Contextual Theory	40
Abbildung 02:	PARTS-Wirksamkeitsmodell	103
Abbildung 03:	Design der PARTS-Studie	121

I. Einleitung

Ein Jugendlicher trifft sich am Nachmittag mit einer Gruppe von Freunden im Garten eines Bekannten. Er ist etwa 14 Jahre alt. Die anderen Kinder sind im gleichen Alter, manche aber auch jünger und erst acht oder neun Jahre alt. Der Jugendliche ist der Anführer der Gruppe. Sie stehen lose um einen alten Baumstumpf herum. „So“, sagt der Anführer. „Jetzt stellt Euch mal vor, der Baumstumpf hier wäre ein Neger. Was würdet Ihr dann mit dem machen?“ Auffordernd blickt der Jugendliche in die Runde. Ohne zu zögern beginnen die Jungen auf den Baumstumpf einzuschlagen und zu treten.

Dieses Szenario ist ebenso erschreckend wie real. Es wurde dem Autor im Verlauf der Interventionsphase von der Klassenlehrerin einer der teilnehmenden Schulklassen berichtet. Dabei ging es um einen ihrer Schüler aus der dritten Klasse, der an dieser Situation beteiligt war. Dies ist kein Einzelfall.

Die politisch motivierte Kriminalität im Bereich Rechtsextremismus steigt in Thüringen seit Jahren langsam aber stetig an (Verfassungsschutzbericht des Thüringer Innenministeriums, 2008). Ebenso wie die Mitgliederzahl in der rechtsextremen Partei NPD (ebenda). Auch auf Bundesebene nehmen die rechtsextrem motivierten Kriminaldelikte zu (Verfassungsschutzbericht des Bundesministeriums des Inneren, 2008). Vorurteile, Stereotype und diskriminierendes Verhalten sind ein Problem, das in allen Teilen unserer Gesellschaft auftaucht. Dabei haben derartige gruppenbezogene soziale Einstellungen auch im gemäßigten, nicht-kriminellen Ausmaß weitreichende Auswirkungen sowohl für Betroffene als auch für Einstellungsträger. So finden sich bei Kindern mit Stereotypen Verzerrungen in der Informationsverarbeitung (Bigler & Liben, 1993) genauso wie in berufsbezogenen Urteilen und Zielen (Bigler, Averhart & Liben, 2003). Auch die akademische Selbstwirksamkeit und der generelle Selbstwert von Betroffenen leiden unter den ihnen entgegengebrachten negativen Vorurteilen (Osborne, 1997). Die Auswirkungen solcher Einstellungen erstrecken sich bis hin zu Präferenzen in der Auswahl von Freundschaften (Martin & Fabes, 2001), Aktivitäts- und Objektpräferenzen (Coker, 1984) und der Ausführung ideologiemotivierten diskriminierenden oder gar gewalttätigen Verhaltens (Frindte, Neumann, Hieber, Knotte & Müller, 2001). Dies alles sind Gründe, die ein wissenschaftlich fundiertes

Eingreifen zur Reduzierung von negativen gruppenbezogenen sozialen Einstellungen und zur Förderung von Toleranz in unserer Gesellschaft sinnvoll und notwendig machen.

Um effektiv Einfluss auf solche sozialen Einstellungen nehmen zu können, sind Kenntnisse über deren Entwicklung und mögliche Einflussfaktoren auf gruppenbezogene soziale Einstellungen notwendig. Betrachtet man den normativen Entwicklungsverlauf, so zeigen sich gruppenbezogene soziale Einstellungen und ihre Auswirkungen bereits in der Kindheit und Jugend (s. Kapitel II.2. Entwicklung von gruppenbezogenen sozialen Einstellungen). Derart früh entwickelte Einstellungen tendieren dazu, sich über die Zeit zu stabilisieren und zu verfestigen (Devine, 1989). Einmal bestehende und gefestigte Einstellungen lassen sich nur schwer beeinflussen und korrigieren (Devine, 1989). Daher ist es sinnvoll vor der Entstehung und Festigung dieser Einstellungen, also in der Kindheit, mit einer Intervention zu beginnen (Hewstone & Brown, 1986). Einen derartigen Ansatz verfolgen Präventionsmaßnahmen. Solche Maßnahmen erweisen sich als besonders wirksam, wenn sie in einem günstigen Zeitfenster der kindlichen Entwicklung ansetzen und dabei die normale Entwicklung stützen sowie in die gewünschte Richtung lenken (Beelmann, Saur, Ziegler, 2010). PARTS wurde dementsprechend für Kinder der dritten Klasse entwickelt, da sich in diesem Alter ein wichtiger Schritt der kognitiven Entwicklung vom voroperationalen Denken hin zum konkret operationalen Denken vollzieht. In Zusammenhang mit dieser kognitiven Entwicklung erfolgt eine Reduktion von Vorurteilen und Stereotypen durch eine neue, differenziertere Wahrnehmung von Mitgliedern der Eigen- und Fremdgruppe (s. Kapitel II.2. Entwicklung von gruppenbezogenen sozialen Einstellungen). Somit wurde für die Durchführung des PARTS-Projektes ein günstiges Zeitfenster der Entwicklung gruppenbezogener sozialer Einstellungen gewählt.

Die Einflussfaktoren, die bei der Entwicklung von gruppenbezogenen sozialen Einstellungen eine Rolle spielen sind vielfältig. Hier sind besonders die kognitive Entwicklung der Kinder (Aboud, 1988), die Entwicklung der sozialen Identität (Nesdale, 2001, 2008) sowie der soziale Kontext dieser Entwicklung stattfindet (Bar-Tal & Teichman, 2005; Teichman & Bar-Tal, 2008) zu nennen. Weiterhin sind die Kontaktmöglichkeiten zu Mitgliedern von Fremdgruppen ein wichtiger Einflussfaktor bei der Einstellungsentwicklung (Raabe & Beelmann, in Druck). Aber natürlich

spielen auch die Eltern-Kind-Beziehung, der Erziehungsstil der Eltern und Lehrer (Adorno, Frenkel-Brunswick, Levinson & Sanford, 1950) und die Lernerfahrungen der Kinder (Bandura, 1977) eine wichtige Rolle bei der Einstellungsentwicklung.

Gruppenbezogene soziale Einstellungen sind kein einheitliches Konstrukt. Im Gegenteil: Vorurteile, Stereotype, diskriminierendes Verhalten oder einfach nur negativer Affekt gegenüber Mitgliedern von Fremdgruppen bilden unterschiedliche Facetten dieser gruppenbezogenen sozialen Einstellungen ab. Dabei können die einzelnen Facetten danach unterschieden werden, ob sie eher kognitive, affektive oder konative Schwerpunkte besitzen. Häufig zeigt sich auch, dass die einzelnen Einstellungsfacetten nur bis zu einem gewissen Grad miteinander zusammenhängen (Dovidio, Brigham, Johnson & Gaertner, 1996; Dovidio, Gaertner, Stewart, Esses, Vergert & Hodson, 2004). Jemand mit Vorurteilen gegenüber einer Gruppe von Menschen, muss also noch längst kein diskriminierendes Verhalten gegenüber dieser Gruppe zeigen. Dies muss bei der Planung einer Intervention zur Prävention von Vorurteilen, Stereotypen und Diskriminierung berücksichtigt werden. Es zeigt sich nämlich, dass bestimmte Interventionen besonders geeignet sind, bestimmte Einstellungsfacetten zu beeinflussen. So eignen sich beispielsweise Kontakt-Interventionen besonders, um affektive Einstellungskomponenten zu verändern, während wissensbasierte Interventionen verstärkt kognitive Einstellungsbereiche beeinflussen (Dovidio et al., 2004). Um eine umfassende Prävention aller Bereiche negativer gruppenbezogener sozialer Einstellungen zu gewährleisten, ist es daher notwendig, ein multimodales Interventionsprogramm zu entwickeln und durchzuführen. Einen solchen multimodalen Ansatz verfolgt PARTS. Das Programm wurde auf der Grundlage von bereits erfolgreich evaluierten Interventionsansätzen entwickelt und basiert auf drei methodischen Säulen: der Förderung individueller sozial-kognitiver Kompetenzen, stellvertretendem Kontakt zu Mitgliedern einer Fremdgruppe und der Vermittlung von interkulturellem Wissen. Dabei steht PARTS als Akronym für *Program* zur Förderung von *Akzeptanz*, *Respekt*, *Toleranz* und *Sozialer Kompetenz*. Dieser Programmname beinhaltet somit die Ziele des Interventionsprogramms. Allerdings ist dies nicht wortwörtlich, sondern eher richtungsgebend zu verstehen. So werden nicht alle Begriffe, die im Programmnamen auftauchen operationalisiert und auf ihre Veränderungen durch das Programm hin geprüft. Vielmehr beschränkt sich die Evaluation von PARTS auf die

Prüfung der Präventionswirkung gegenüber gruppenbezogenen sozialen Einstellungen sowie der Reduktion von Intoleranz und der Förderung von Akzeptanz und sozialer Kompetenz. Der Name PARTS soll vor allem den positiven Fokus des Programms aufzeigen sowie für alle Beteiligten griffig und gut verknüpfbar sein.

Ziel der vorliegenden Arbeit ist es, wissenschaftlich fundierte Aussagen über die Umsetzbarkeit und Wirksamkeit des neu entwickelten Präventionsprogramms PARTS treffen zu können. Die hierfür notwendigen Informationen werden in den einzelnen Kapiteln dieser Arbeit dargelegt. So werden zunächst die für eine Intervention erforderlichen Grundlagen erläutert: die Begrifflichkeiten, der Verlauf der normativen Einstellungsentwicklung sowie die einschlägigsten Theorien zur Erklärung der Entwicklung gruppenbezogener sozialer Einstellungen. Weiterhin werden die für diese Arbeit zentralen Interventionsstudien aus den Bereichen Stellvertretender-Kontakt-Interventionen, wissensbasierte Interventionen und Förderung individueller Kompetenzen ausführlich referiert. All das sind wichtige Grundlagen, um ein wissenschaftlich fundiertes Interventionsprogramm zu entwickeln und evaluieren zu können. Im Methodenteil werden daraufhin das Interventionsprogramm, die Charakteristika der Stichprobe, das Design der Studie, der Ablauf der Untersuchungen sowie die verwendeten Messinstrumente beschrieben. Weiterhin werden die Zielsetzungen des Interventionsprogramms und deren Operationalisierung dargestellt. Schließlich werden die Ergebnisse der Implementationsanalyse und der Evaluation des Programms ausführlich berichtet und vor dem Hintergrund des aktuellen Forschungsstandes diskutiert. Somit ermöglicht diese Arbeit, gesicherte Aussagen über die Umsetzbarkeit und Wirksamkeit von PARTS treffen zu können.

II. Theorie

II.1. Grundlagen

In der vorliegenden Arbeit geht es um die Evaluation des Präventionsprogramms PARTS zur Förderung von positiven interpersonalen sozialen Einstellungen wie Toleranz, sowie zur Reduktion von negativen gruppenbezogenen sozialen Einstellungen wie Stereotypen, Vorurteilen und Diskriminierung in Intergruppenbeziehungen. Um die Wirksamkeit von PARTS wissenschaftlich prüfen zu können, müssen klare Zielkriterien erstellt werden. Hierzu müssen zunächst allerdings einige notwendige Grundlagen erläutert und Begriffe definiert werden. Hier werden die zentralen Begriffe der Arbeit definiert und voneinander abgegrenzt.

II.1.1. Gruppenbezogene soziale Einstellungen

Unter dem Oberbegriff der gruppenbezogenen sozialen Einstellungen lassen sich verschiedene Formen sozialer Einstellungen zusammenfassen. Einstellungen wie Vorurteile oder Stereotype, aber auch damit zusammenhängende Verhaltensweisen wie Diskriminierung besitzen zwar alle ihre eigene spezifische Bedeutung, verfügen jedoch auch über Gemeinsamkeiten. Da es in der in der vorliegenden Arbeit um die Evaluation eines Programms zur Prävention von Vorurteilen, Stereotypen und Diskriminierung geht, scheint es zunächst sinnvoll und vereinfachend diese verschiedenen Konstrukte unter einem Sammelbegriff zu subsummieren. Daher wird im Folgenden häufig von gruppenbezogenen sozialen Einstellungen gesprochen, wenn Vorurteile, Stereotype und Diskriminierung gemeint sind. Um aber trotzdem eine differenzierte Analyse der Interventionseffekte zu ermöglichen, werden diese gruppenbezogenen sozialen Einstellungen nach unterschiedlichen Schwerpunkten spezifiziert. Dabei gibt es einen affektiven, einen kognitiven und einen konativen Schwerpunkt. Unabhängig davon werden interpersonale Einstellungen wie Toleranz und allgemeines Sozialverhalten getrennt betrachtet.

Doch zunächst werden die einzelnen Einstellungsformen von gruppenbezogenen sozialen Einstellungen vorgestellt und einem der drei Schwerpunkte zugeordnet.

Vorurteile

Gordon W. Allport beschreibt in seinem Buch „The Nature of Prejudice“ (1954) Vorurteile wie folgt:

„... prejudice is an antipathy based upon a faulty and inflexible generalization. It may be felt or expressed. It may be directed toward a group as a whole, or toward an individual because he is a member of that group“ (S. 9).

Allport nennt drei essentielle Bestandteile von Vorurteilen: a) Feindseligkeit und Ablehnung, b) der Hinweisreiz, der die Feindseligkeit auslöst, liegt in der Gruppenzugehörigkeit des „Opfers“ und c) Widerstand gegenüber neuem Wissen, welches den Vorurteilen widerspricht (Allport, 1954). Weiterhin unterscheidet Allport zwischen zwei Komponenten von ethnischen Vorurteilen, von denen die erste eine affektive Bewertung einer Gruppe beinhaltet, während die zweite Attribute umfasst, die dieser Gruppe zugeschrieben werden.

Nachfolgend fokussierten andere Autoren auf bestimmte Aspekte des Vorurteilsbegriffs. So beschreibt Ehrlich (1973) Vorurteile als ungünstige Einstellungen gegenüber Personen auf Grund der Zugehörigkeit zu einer Gruppe. Er legt den Fokus besonders auf die affektive Komponente der Einstellung und weniger auf die Zuschreibung von Eigenschaften, den kognitiven Anteil. Fishbein (2002) und Milner (1983) konstatieren hierzu, dass diese Affekte häufig irrational und unbegründet seien.

Vorurteile sind außerdem sehr stabile Einstellungen, die sich, sobald sie einmal vorhanden sind, nicht mehr leicht verändern lassen (Allport, 1954; Devine, 1989). Sie erweisen sich sowohl als zeitlich stabil als auch situationsübergreifend aktiv (Aboud, 1988).

Nesdale (2004) grenzt den Vorurteilsbegriff von der puren Bevorzugung der eigenen Gruppe gegenüber anderen Gruppen ab und betont in seiner Arbeit, dass Vorurteile ausschließlich als negative Einstellungen („hate/dislike“ also Hass bzw. Abneigung) bzw. Verhalten („derogation“ also Abwertung) gegenüber Fremdgruppenmitgliedern zu betrachten seien. Diese Einstellungen entwickeln sich nicht bei jedem Menschen, sondern nur bei Personen, die sich mit Gruppen identifizieren, welche Vorurteile pflegen. Die Eigengruppenpräferenz, also die

Bevorzugung der Eigengruppe gegenüber fremden Gruppen allein, sieht Nesdale als Teil einer normalen Entwicklung und lässt es nicht als Vorurteil gelten. Damit kritisiert er diverse Studien, die auch unterschiedliche Positivbewertungen von Eigen- und Fremdgruppe als Indikator für Vorurteile gewertet haben. Somit eröffnet Nesdale eine weitere Unterscheidungsdimension von gruppenbezogenen sozialen Einstellungen: Eigengruppenbevorzugung versus Fremdgruppenabwertung. Insgesamt lässt sich damit festhalten, dass der Begriff des Vorurteils durch folgende Merkmale gekennzeichnet ist:

- Eine übergeneralisierte negative Einstellung/Bewertung einer Person auf Grund deren Zugehörigkeit zu einer Gruppe.
- Diese Einstellung beinhaltet vor allem eine affektive Komponente, aber auch einen kognitiven Anteil und geht mit einer Verhaltensdisposition einher.
- Vorurteile sind relativ stabil (zeitlich und situativ) und resistent gegen Veränderungen.
- Vorurteile können auf zwei Dimensionen erfasst werden: Die Eigengruppenbevorzugung und die Fremdgruppenabwertung.

Somit lässt sich der Begriff des Vorurteils als gruppenbezogene soziale Einstellung mit affektivem Schwerpunkt einordnen.

Stereotype

„Stereotype sind eine kognitive Struktur, die Wissen, Überzeugungen und Erwartungen gegenüber einer sozialen Gruppe enthält“ (Hamilton und Trolie, 1986, S. 133). Dabei kann sich der Inhalt von Stereotypen im Laufe der Zeit verändern (Aarim-Herlot, 2003; Ignatiev, 1995). Nach Brown (1986) sind Stereotype die geteilte Konzeption der Charakteristika einer Gruppe.

Milner (1983) beschreibt Stereotype als Übergeneralisierung der Charakteristika einer Gruppe, die meist unerwünscht ist, und dazu dient, die Unterschiede zwischen Gruppen zu übertreiben. Die Entwicklung und Anwendung von Stereotypen als Kategorie von Überzeugungen über Gruppen von Menschen hat aber auch eine normale und notwendig-adaptive Funktion (Fishbein, 2002).

Stereotype können sowohl positive als auch negative Inhalte haben. Bei negativen Stereotypen ist die Abgrenzung zum Begriff des Vorurteils zwar schwierig, aber möglich. Stereotype müssen nicht unbedingt irrational sein, wie Fishbein (2002)

und Milner (1983) es von Vorurteilen verlangen. Sie haben auch nicht zwingend eine affektive Komponente, noch beinhalten sie eine Verhaltensdisposition. Somit sind Stereotype eine überwiegend kognitive Einstellung oder Überzeugung. Allerdings lassen sich trotzdem Korrelationen zwischen Stereotypen und Vorurteilen finden ($r = .25$, Dovidio, Brigham, Johnson & Gaertner, 1996). Beide Konstrukte (Vorurteile und Stereotype) sind also miteinander verknüpft, verfügen aber trotzdem über genügend spezifische Varianz, um die Verwendung unterschiedlicher Begriffe zu rechtfertigen.

Somit lässt sich festhalten, dass Stereotype folgende Merkmale aufweisen:

- Stereotype sind eine kognitive Struktur, die übergeneralisierendes Wissen, Überzeugungen und Erwartungen gegenüber einer sozialen Gruppe beinhalten.
- Stereotype beinhalten keine affektive Komponente und auch keine Verhaltensdisposition und lassen sich somit vom Begriff des Vorurteils abgrenzen.
- Die Inhalte von Stereotypen sind sozial geteilt, abhängig vom aktuellen Zeitgeist und können in ihrer Valenz sowohl positiv als auch negativ oder neutral sein.

Somit lässt sich der Begriff des Stereotyps als gruppenbezogene soziale Einstellung mit kognitivem Schwerpunkt einordnen.

Diskriminierung

Wie oben bereits beschrieben, beinhalten Vorurteile drei Komponenten: Eine kognitive, eine affektive und eine behaviorale Komponente. Dabei besteht die behaviorale Komponente allerdings nur aus der Disposition, sich negativ gegenüber Mitgliedern der Zielgruppe der Vorurteile zu verhalten und nicht aus dem Verhalten selbst (Ehrlich, 1973). Das Verhalten selbst wird häufig als Diskriminierung bezeichnet. Diskriminierung wird dabei als schädliche Handlung gegenüber anderen auf Grund deren Mitgliedschaft in einer bestimmten Gruppe verstanden (Fishbein, 2002; Marger, 1991; Allport, 1954).

Diskriminierendes Verhalten kann unterschiedlich stark ausgeprägt sein. Angefangen beim Ignorieren des Opfers oder heimlichen Lästereien bis hin zu Mord und Totschlag gibt es eine große Varianz in der Intensität von Diskriminierung. Bei

Kindern äußert sich Diskriminierung meist durch Vermeidung, Grobheit, Beschimpfungen oder seltener durch Prügeleien (Fishbein, 2002).

Häufig wird vermutet, dass Diskriminierung auf Vorurteilen beruht. Allerdings kann es auch diskriminierendes Verhalten ohne Vorurteile geben, wenn das Individuum durch äußere Zwänge zu diesem Verhalten veranlasst wird. Weiterhin scheint es eher so zu sein, dass Diskriminierung und Vorurteile sich gegenseitig verstärken und somit eine wechselseitige Beziehung zwischen beiden Konstrukten besteht (Frederickson & Knobel, 1980). Somit kann der Begriff der Diskriminierung auch als gruppenbezogene soziale Einstellung mit konativem Schwerpunkt verstanden werden.

II.1.2. Toleranz

Der Begriff der Toleranz lässt sich nicht unter den Oberbegriff der gruppenbezogenen sozialen Einstellungen einordnen. Toleranz ist vielmehr eine allgemeine interpersonale Einstellung, die sich auch auf gruppenbezogene Kontexte anwenden lässt. Der Begriff der Toleranz ist bisher allerdings weniger stark beachtet und erforscht worden als die Begriffe Vorurteile, Stereotype und Diskriminierung. Eine Einordnung dieses Begriffs in einen grundsätzlichen definitorischen Rahmen bieten Van Quaquebeke, Henrich und Eckloff (2007).

Toleranz ist demnach eine Art von Einstellung, ähnlich auch Akzeptanz oder Respekt. Diese drei Begriffe sind sich inhaltlich relativ nah und müssen daher gut voneinander abgegrenzt werden. Van Quaquebeke und Kollegen unterscheiden Toleranz, Akzeptanz und würdigenden Respekt¹ in Bezug auf den Sachverhalt auf den sich diese Einstellungen beziehen: Toleranz bezieht sich auf die Präsenz eines Objekts in der Umgebung des Subjekts. Akzeptanz bezieht sich auf die Mitgliedschaft in der Gruppe des Subjekts und würdigender Respekt bezieht sich auf den Einfluss des Objekts auf das Subjekt.

Die Toleranz des Subjekts legitimiert das Objekt, sich im Wahrnehmungsfeld des Subjekts aufzuhalten. Um von Toleranz gegenüber einem Objekt sprechen zu können, muss dieses in irgendeiner Weise von einer Norm abweichen. Diese Abweichung wird vom Subjekt „ertragen“, auch wenn es sie missbilligt und etwas dagegen unternehmen könnte.

¹ Im Vergleich zum würdigenden Respekt, setzten die Autoren den wahrnehmenden Respekt (s.u.).

Unter Akzeptanz verstehen die Autoren die Aufnahme des Objekts in die Gruppe des Subjekts.

Der würdigende Respekt bezieht sich auf die Bewunderung von exzellenten Leistungen oder Fähigkeiten. Die Autoren vermuten aber als eigentliche Kernkomponente die Gestattung der Einflussnahme des respektierten Objekts auf das Subjekt in dem Sinne, dass das Objekt als Vorbild für das Subjekt dient, oder dass das Subjekt den Lehren des Objekts folgt. Das Gegenteil dieses würdigenden Respekts ist die Reaktanz, in der sich das Subjekt der Einflussnahme des Objekts widersetzt.

All diesen Begriffen ist dabei gemein, dass diese Einstellungen nur unter bestimmten Bedingungen auftreten. Sie sind also nicht unbedingt, sondern benötigen die Salienz des entsprechenden Themas um aufzutreten. Weiterhin ist eine Voraussetzung für die Anwendung dieser Eigenschaften, dass die entsprechenden Themen der Legitimität, Gruppenmitgliedschaft oder Einflussnahme überhaupt für das Subjekt salient werden. Ohne die Wahrnehmung dieser Themen kann das Subjekt auch keine entsprechende Einstellung bzw. kein bewusstes Verhalten dazu zeigen.

Die Autoren sehen im Gegensatz zu den oben beschriebenen Begriffen noch eine andere Form des Respekts: den wahrnehmenden Respekt. Dieser ist eine prozessorientierte Einstellung und beinhaltet die Wahrnehmung und Behandlung des Gegenübers als gleichwertig. Er bezieht sich damit vor allem auf die Art und Weise der Interaktion und weniger auf deren Ausgang. Dieser wahrnehmende Respekt ist im Gegensatz zu den drei anderen genannten Einstellungen unbedingt und nicht themenbezogen, sondern eher eine allgemeine Grundhaltung anderen Menschen gegenüber.

Die Autoren unterteilen nun die Begriffe Toleranz, Akzeptanz und würdigender Respekt zusätzlich noch in eine respektvolle und eine respektlose Form (s. Tabelle 01). Diese Unterteilung bezieht sich auf den wahrnehmenden Respekt und damit auf die Grundhaltung in der Einstellung und Interaktion des Subjekts gegenüber dem Objekt.

Tabelle 01

Abgrenzung der Begriffe Toleranz, Akzeptanz und zwei Formen des Respekts

Begriff	Sachverhalt	Einstellungsart	Bedingungen
Toleranz	Legitimität des Objekts	ergebnisorientiert	Salienz
Akzeptanz	Aufnahme als Gruppenmitglied	ergebnisorientiert	Salienz
Würdiger Respekt	Gestattung der Einflussnahme	ergebnisorientiert	Salienz
Wahrnehmender Respekt	Wahrnehmung und Behandlung des Objekts als gleichwertig	prozessorientiert	

Anmerkungen.

Nicht alle der oben definierten Begriffe werden bei der Evaluation von PARTS operationalisiert und betrachtet, auch wenn sie im Namen PARTS vorkommen. Hier mangelt es in der Forschung vor allem an hochwertigen Instrumenten zur Messung solcher interpersonaler Einstellungen. Die Begriffe der Toleranz bzw. Intoleranz und der Akzeptanz spielen jedoch auch für die Evaluation eine wichtige Rolle. Hier wird eine selbstentwickelte Skala zur Messung der beiden Konstrukte verwendet (s. Kapitel III.6. Messinstrumente). Die übrigen Begriffsdefinitionen dienen eher der Abgrenzung der umgangssprachlich ähnlich verwendeten Begriffe voneinander.

II.1.3. Ausdrucksformen gruppenbezogener sozialer Einstellungen

Zusammenfassend gilt: Gruppenbezogene soziale Einstellungen können in unterschiedlichen Situationen auftreten und dabei abhängig von der Situation und der Person unterschiedliche Ausdrucksformen annehmen. So kann zum einen nach den strukturellen Schwerpunkten dieser Einstellungen unterschieden werden. Es kann sich um übergeneralisierte kognitive Eigenschaftszuschreibungen handeln, aber auch um affektiv aufgeladene negative Überzeugungen oder gar entsprechendes negatives Verhalten gegenüber Mitgliedern von Fremdgruppen (s.o. Stereotype, Vorurteile und Diskriminierung).

Auch das Zentrum der gruppenbezogenen sozialen Einstellung dient als Unterscheidungskriterium (Allport, 1954; Brewer, 1999). Wenn die Eigengruppe des Subjekts im Zentrum der Einstellungen liegt, spricht man von der Eigengruppenpräferenz. Hier wird die Eigengruppe in unterschiedlichen Bereichen wie Eigenschaftszuschreibungen, Ressourcenaufteilungen oder prosozialem Verhalten im Vergleich zu Fremdgruppen bevorzugt behandelt (Mullen, Brown & Smith, 1992; Messick & Mackie, 1989).

Demgegenüber steht die Fremdgruppenabwertung, bei der die Fremdgruppe im Zentrum der Aufmerksamkeit steht (Nesdale, 2008; Brewer, 1999). Hier werden Mitglieder von Fremdgruppen nicht nur weniger positiv als die Eigengruppe, sondern direkt negativ bewertet, behandelt oder ganz aus sozialen Situationen ausgeschlossen.

Diese Unterscheidung ist sowohl von theoretischem als auch von praktischem Interesse. Zum einen werden unterschiedliche zugrunde liegende Prozesse zwischen Eigengruppenpräferenz und Fremdgruppenabwertung vermutet (Nesdale, 2008; Brewer, 1999). Somit ergeben sich auch unterschiedliche Interventionsansätze zur Beeinflussung der beiden Einstellungsformen. Zum anderen haben diese beiden Einstellungsformen natürlich auch unterschiedliche praktische Implikationen für die Einstellungsträger und die Einstellungsziele.

Die zentralen Begriffe und Konstrukte für die Evaluation von PARTS sind in diesem Kapitel klar definiert und voneinander abgegrenzt worden. Auf dieser Basis ist das Setzen von klaren Zielkriterien für das Interventionsprogramm möglich. Doch bevor die Zielkriterien von PARTS festgelegt werden, sind Informationen über die normative Entwicklung von Einstellungen bei Kindern notwendig. Diese werden im folgenden Kapitel dargestellt.

II.2. Die Entwicklung gruppenbezogener sozialer Einstellungen

Nur im Vergleich zur normativen Entwicklung von gruppenbezogenen sozialen Einstellungen können Zielkriterien, die zwischen Erfolg oder Misserfolg der Intervention trennen, entwickelt werden. Weiterhin muss der normative Entwicklungsverlauf bekannt sein, um eine entwicklungspsychologisch fundierte Intervention entwickeln zu können. Nur so kann man den optimalen Zeitpunkt für eine Intervention identifizieren und das Programm auf den Entwicklungsstand der Kinder abstimmen. Daher wird im ersten Abschnitt dieses Kapitels der generelle Entwicklungsverlauf von gruppenbezogenen sozialen Einstellungen auf der Basis der aktuellen Forschungsliteratur dargestellt.

Die ontogenetische Entwicklung von Vorurteilen ist jedoch nicht nur relevant, um Zielkriterien und den idealen Zeitpunkt für Interventionen zu identifizieren, sondern auch um mögliche Interventionsansätze auf Grundlage der ablaufenden Entwicklungsprozesse herauszuarbeiten. Daher werden neben dem Entwicklungsverlauf auch unterschiedliche Erklärungsmodelle und Theorien für die Entwicklung von gruppenbezogenen sozialen Einstellungen im zweiten Abschnitt des Kapitels beschrieben.

II.2.1. Der Entwicklungsverlauf von Vorurteilen und Stereotypen

Leider existieren bis dato nur sehr wenige längsschnittliche Studien zur Entwicklung von Vorurteilen bei Kindern (vgl. Raabe & Beelmann, in Druck). Dennoch lässt sich mit den Ergebnissen der vorliegenden Längsschnittstudien sowie den zusammengesetzten Ergebnissen verschiedener querschnittlicher Untersuchungen ein charakteristischer normativer Entwicklungsverlauf von Vorurteilen und Stereotypen bei Kindern nachzeichnen.

Gruppenbezogene soziale Einstellungen treten bereits sehr früh in der Kindheit auf. Aboud (1988), Bar-Tal (1996), als auch Nesdale (2001) können kulturübergreifend zeigen, dass die ersten Vorstufen solcher Einstellungen, wie die Unterscheidung zwischen Personen verschiedener Ethnien und ein damit verbundener Affekt, im Alter zwischen drei und vier Jahren erkennbar werden. Diese frühen Einstellungen finden sich, ohne dass die Kinder dabei genaueres über diese Gruppen wissen oder sie einwandfrei definieren können (Bar-Tal, 1996). Im weiteren

Entwicklungsverlauf steigen Vorurteile und Stereotype bis zu einem Alter von etwa sieben Jahren an. Kinder im Alter von sieben Jahren können sich selbst zuverlässig ihrer ethnischen Gruppe zuordnen, was eine wichtige Voraussetzung für die Vorurteilsentwicklung darstellt. Weiterhin zeigen sie verstärkte Eigengruppenpräferenz und Fremdgruppenabwertung bei Eigenschaftszuschreibungen (Aboud, 1988, Nesdale, 2001). Im Altersbereich ab sieben Jahren sinkt die Intensität dieser Einstellungen dann wieder etwas ab (Aboud, 1988, Nesdale 2000, Aboud & Amato, 2001). So konnten Doyle und Aboud (1995) in einer der wenigen längsschnittlichen Studien zeigen, dass ältere Kinder (neun Jahre alt) weniger vorurteilsbelastet sind als jüngere (sechs Jahre alt). Dabei verringert sich die vorherrschende positive Bewertung der Eigengruppe. Die negative Bewertung der Fremdgruppe allerdings nicht, sondern die Zuschreibung negativer Eigenschaften zur Eigengruppe und positiver Attribute zur Fremdgruppe, der so genannte Counterbias, steigt an. Dies legt nach Aboud (1988, 2008) den Schluss einer mit dem Alter zunehmend differenzierteren individualisierten Wahrnehmung von Personen unabhängig von der Gruppenzugehörigkeit nahe, da den zu beurteilenden Personen sowohl positive als auch negative Eigenschaften zugeordnet wurden.

Im weiteren Entwicklungsverlauf wird angenommen, dass Vorurteile nach dem Absinken in der mittleren Kindheit bis zur Adoleszenz langsam wieder ansteigen (Nesdale, 2000).

Der hier nachgezeichnete Entwicklungsverlauf von Vorurteilen scheint jedoch nicht für alle Kinder Gültigkeit zu besitzen und sollte nicht bedenkenlos generalisiert werden (Raabe und Beelmann, in Druck). So konnten Branch und Newcombe (1986) zeigen, dass die Vorurteilsentwicklung bei Minoritätskindern verzögert abläuft, da diese in der frühen Kindheit zunächst die Majoritätsgruppe, also nicht die Eigengruppe präferieren. Auch die Entwicklung von nationalen Stereotypen und Vorurteilen, die nicht an äußerlichen Merkmalen wie der Hautfarbe der Gruppenmitglieder festgemacht werden können, scheint anders und verzögert zu verlaufen (Rutland, 1999).

Insgesamt ist die Ausprägung der Vorurteile im Kindesalter eher gering, lässt sich jedoch im Vergleich zwischen Eigengruppe und Fremdgruppe als relative Einstellungsverzerrung einzig auf Grund der Gruppenzugehörigkeit finden (Aboud, 2008).

Ungeachtet dessen konnten Raabe und Beelmann (in Druck) in ihrer Metaanalyse zur Entwicklung von Vorurteilen aus 145 Arbeiten zu diesem Thema den oben geschilderten Verlauf, zumindest für Vorurteile gegenüber ethnischen und nationalen Gruppen (90% der Studien), bestätigen. Sie konnten zeigen, dass Vorurteile von der frühen bis zur mittleren Kindheit ansteigen, im Alter von sechs bis sieben Jahren ihren Höhepunkt erreichen, um dann in der späten Kindheit (acht bis zehn Jahren) abzufallen. Faktoren die für den hier nachgezeichneten Entwicklungsverlauf eine Rolle zu spielen scheinen, sind: die sozial-kognitive Entwicklung, die Anpassung an und Sensitivität für soziale Normen (sind Vorurteile in der Gesellschaft gegenüber dieser Gruppe geduldet oder nicht), sowie die meist explizite Messung von Vorurteilen in den beteiligten Arbeiten (soziale Erwünschtheit).

Da Interventionsprogramme erfolgreicher verlaufen, wenn sie der normativen Entwicklung zeitlich und inhaltlich angepasst sind (Beelmann, 2000, Sameroff & Fiese, 2000) sollten Interventionen zur Prävention von gefestigten gruppenbezogenen sozialen Einstellungen in der mittleren Kindheit, zwischen acht und neun Jahren, einsetzen. In diesem Zeitraum scheinen gruppenbezogene soziale Einstellungen einer normativen Entwicklung und Differenzierung zu unterliegen. Daher sollte dabei unter anderem auf die sozial-kognitive Entwicklung der Kinder in diesem Alter eingegangen werden und eine differenzierte Betrachtung von Personen gefördert werden.

II.2.2. Entwicklungstheorien

Um eine effektive Intervention entwickeln zu können, muss klar sein, welche Prozesse für die Entwicklung von gruppenbezogenen sozialen Einstellungen wie Vorurteile und Stereotype relevant sind. Diese Prozesse bieten Ansatzpunkte für mögliche Interventionen.

Die entwicklungspsychologischen Prozesse, die bei der Ausbildung von Stereotypen und Vorurteilen eine Rolle spielen, sind allerdings längst nicht vollständig aufgeklärt. Unterschiedliche Perspektiven sowie eine bisher noch unbefriedigende Datenlage ergeben ein komplexes und teilweise lückenhaftes Bild (Beelmann, Saur & Schulze, 2006). Im Folgenden wird zunächst ein kurzer Überblick über die Theorienlandschaft in diesem Forschungsbereich gegeben. Dann werden drei aktuelle, für diese Arbeit zentrale Entwicklungstheorien vorgestellt.

Die Entwicklung von Vorurteilen wurde in den dreißiger und vierziger Jahren das erste Mal untersucht. Damalige Studien hatten noch keine theoretische Basis und befassten sich hauptsächlich mit der ethnischen Präferenz von Kindern (Horowitz, 1936, Horowitz & Horowitz, 1938; Clark & Clark, 1947). In den fünfziger Jahren wurden dann die ersten Theorien zur Entwicklung von Vorurteilen erarbeitet.

So entwickelte die Arbeitsgruppe um Theodor W. Adorno ihre Theorie der autoritären Persönlichkeit (Adorno, Frenkel-Brunswik, Levinson & Sanford, 1950). Diese Theorie basiert auf einem psychoanalytischen Persönlichkeitsmodell und postuliert, dass Vorurteile mit einem besonderen Typ, der autoritären Persönlichkeit einhergehen. Diese autoritäre Persönlichkeit entwickelt sich vor allem durch bestimmte Erziehungspraktiken, wie ein sehr kühles Familienklima und einen fordernden, kontrollierenden, wenig unterstützenden Erziehungsstil.

Eine andere einflussreiche Theorie zur Vorurteilsentwicklung basiert auf der Sozialisationstheorie und geht, im Gegensatz zu den Forschern der Frankfurter Schule um Adorno, davon aus, dass Menschen Vorurteile entwickeln, indem sie diese von ihrer Umwelt erlernen. Dazu konstatieren sie, dass das Kind tabula rasa, also vorurteilsfrei zur Welt kommt. Im Laufe seiner Sozialisation erlernt das Kind die Einstellungen, die ihm in seiner Umwelt begegnen (Allport, 1954). Dabei spielen Mechanismen wie das soziale Lernen oder auch Modelllernen nach Bandura (1977) eine wichtige Rolle (Lipscomb, McAllister & Bregman 1985). Zusätzlich zu dem Faktor der Nachahmung in vorurteilsbehafteten Familien, wird dieses, am Modell gelernte Verhalten der Kinder, von den Eltern bewusst oder unbewusst durch positive Reaktionen verstärkt (Carter & Rice, 1997).

Neben diesen älteren und eher unspezifischen Erklärungsansätzen für die Entwicklung von Vorurteilen gibt es seit den achtziger Jahren aktuellere Ansätze, die sich speziell mit der Entwicklung gruppenbezogener sozialer Einstellungen befassen. Aus einer Reihe von Theorien werden im Folgenden drei für die Entwicklung und Evaluation von PARTS relevante Theorien vorgestellt.

Aktuelle Entwicklungstheorien

Die drei hier beschriebenen aktuellen Entwicklungstheorien haben für die vorliegende Arbeit, speziell für die Konzeption des Interventionsprogramms, besondere Bedeutung. Zuerst wird auf die Social-Cognitive Developmental Theory of Prejudice (SCT, Aboud, 1988, 2008) eingegangen. Diese Theorie verknüpft die

Entwicklung von Vorurteilen mit der Entwicklung sozial-kognitiver Fähigkeiten der Kinder. Zweitens wird die Social-Identity-Developmental Theory nach Nesdale (SIDT, 2004) vorgestellt. Nesdale legt seinen Fokus bei der Vorurteilsentwicklung auf motivationale Prozesse im Zuge der Identitätsentwicklung des Kindes. Als drittes wird die Integrative Developmental Contextual Theory von Daniel Bar-Tal und Yona Teichman (IDCT, 2006) eingeführt, welche verschiedene Einflussfaktoren und Prozesse in ihrem Modell integriert und dabei besonders den sozialen Kontext der Entwicklung berücksichtigt.

Social-Cognitive Developmental Theory of Prejudice

Frances Aboud hat mit ihrer Arbeit „Children and Prejudice“ (1988) einen wichtigen Baustein für die Erforschung der kindlichen Vorurteilsentwicklung geliefert. Ihre Social-Cognitive Developmental Theory of Prejudice (SCT) will den Erwerb und die altersabhängigen Veränderungen von Vorurteilen bei Kindern erklären. Dazu orientiert sie sich unter anderem am kognitiven Entwicklungsverlauf, wie er von Piaget beschrieben wurde. Gleichzeitig geht Aboud auf empirische Ergebnisse zum Verlauf von Vorurteilen über die Zeit ein. Ihre zentrale These ist dabei, dass die Entwicklung von Vorurteilen bei Kindern abhängig von deren sozial-kognitiver Entwicklung ist.

Sie betrachtet die Vorurteilsentwicklung als Ergebnis von zwei sich überlappenden Entwicklungssequenzen. Die erste Sequenz betrifft die für das Kind relevante Erfahrungsebene, somit die dominante psychologische Prozessebene.

Sequenz der Prozessebene: Affektiv – Perzeptiv – Kognitiv.

Die erste Phase dieser Sequenz wird von affektiv-emotionalen Prozessen dominiert. In dieser Zeit sind besonders die emotionale Bindung, die Angst vor dem Unbekannten oder die persönlichen Vorlieben für ein Kind bedeutsam (Aboud, 2008). Die zweite Phase wird von perzeptiven Prozessen bestimmt. Die Identifizierung und Beobachtung von Hinweisreizen, die am Kind selbst oder an anderen Personen wahrgenommen werden, ermöglichen es dem Kind, Zuordnungen und Kategorisierungen durchzuführen. Ähnlichkeiten und Unähnlichkeiten werden stärker fokussiert und sind bedeutsam für die Kategorisierung und Bewertung von Objekten. Ähnliches wird positiv bewertet, Unähnliches negativ. Ab einem Alter von sieben Jahren werden schließlich die kognitiven Prozesse immer differenzierter und damit

auch dominanter. Abstrakte oder interne Qualitäten von Personen werden in Bewertungen einbezogen und auch unterschiedliche Perspektiven können langsam miteinander in Beziehung gesetzt werden.

Die zweite Sequenz verläuft über den Fokus der Aufmerksamkeit des Kindes.

Sequenz des Aufmerksamkeitsfokus: Selbst – Gruppe – Individuum.

Der Aufmerksamkeitsfokus liegt zunächst auf dem Selbst des Kindes. Es ist egozentrisch. Im nächsten Schritt wird der Fokus auf die Gruppe gelegt. Das Kind ist soziozentrisch und achtet sehr auf Gruppen und Gruppenstereotype. Dabei kontrastiert es die Eigen- und Fremdgruppen sehr stark. Kinder haben in dieser Phase eine positive Einstellung zu ihrer Gruppe und können andere Gruppen meist nicht genauso positiv bewerten wie ihre eigene. Hier zeigen die Kinder die stärksten Vorurteile. Schließlich erreicht das Kind eine individuumszentrierte Perspektive mit der es Personen nach ihren individuellen Qualitäten beurteilt und weniger nach den Qualitäten der Gruppe. Damit minimiert sich der Einfluss kategorialer Informationen zu Gunsten individueller Qualitäten in der Personenbeurteilung. Die Vorurteile der Kinder nehmen ab.

Empirisch erfährt die Social Cognitive Developmental Theory of Prejudice einige Unterstützung. So kann mittels zahlreicher Studien der von Aboud beschriebene Entwicklungsverlauf von Vorurteilen bestätigt werden (s.o. Abschnitt Entwicklungsverlauf von Vorurteilen und Stereotypen), auch wenn die Generalisierbarkeit dieses Entwicklungsverlaufs Einschränkungen in Bezug auf die Gruppenzugehörigkeit (Verlauf gilt nur für Kinder der Majoritätsgruppe) (Aboud, 2008), die betreffende ethnische Zielgruppe (Verlauf gilt für Vorurteile gegenüber visuell unterscheidbaren ethnischen Zielgruppen) (Rutland, 1999) sowie den sozialen Kontext (Verlauf gilt nicht in Kontexten dauerhafter Konflikte) (Bar-Tal & Teichman, 2006) erfährt.

Die als Mediator des Entwicklungsverlaufs angenommene Sequenz der dominanten Prozessebene (Affektiv – Perzeptiv – Kognitiv) lässt sich ebenfalls an verschiedenen Punkten empirisch stützen. So finden Bigler und Liben (1993) mit der Fähigkeit zur multiplen Klassifikation der Kinder einen Indikator für den Wechsel zwischen perzeptiver und kognitiver Prozessebene. Diese Fähigkeit bzw. ein Mangel daran stellt ebenfalls ein Korrelat von Vorurteilen dar (Aboud, 2003). Auch konnte

gezeigt werden, dass eine kognitive Differenzierung der Fremdgruppe, wie z.B. durch stereotyp-kontroverse Diskussionen über Fremdgruppenmitglieder, zu einer Senkung von Vorurteilen führt (Aboud & Doyle, 1996b)

Auch die zweite sozial-kognitive Entwicklungssequenz des Aufmerksamkeitsfokus vom selbst- bis hin zum individuumzentrierten Kind und deren Zusammenhang mit gruppenbezogenen sozialen Einstellungen kann zumindest an einzelnen Punkten der Sequenz empirisch gestützt werden. So zeigen soziozentrische Kinder eine starke Bindung an die eigene Gruppe und verstärkten intergruppalen Verzerrungen (Aboud, 2003), während eine Abnahme von Vorurteilen und ein Anstieg des Counterbias mit der neu erworbenen Fähigkeit zur Reconciliation (Akzeptanz der Meinung anderer) einhergehen (Doyle & Aboud, 1995, Aboud, 2002). Auch hier geht eine individuelle differenzierte Betrachtung von Fremdgruppenmitgliedern mit einer Verringerung von Vorurteilen einher (Clark, Hocevar & Dembo., 1980; Katz, Sohn & Zalk, 1975).

Als Kritik an der SCT werden immer wieder ähnliche Punkte angeführt. Zunächst geht es dabei häufig um die Frage der Definition von Vorurteilen. Dabei konstatiert beispielsweise Nesdale (2004), dass Kinder in diesem Alter (fünf bis sieben Jahre) gar keine Vorurteile haben, weil sie nur sehr geringe negative Beurteilungen aussprechen. Aboud hingegen betrachtet relative Maße (Biasmaße) welche die Bewertungen verschiedener Gruppen miteinander vergleicht und vorgefundene Differenzen als Vorurteile interpretiert. Dies erscheint auch sinnvoll, wenn man daran interessiert ist, gruppenbezogene soziale Einstellungen in dieser Altersgruppe zu untersuchen. Als weiterer – diesmal berechtigter – Kritikpunkt wird genannt, dass die meisten Erwachsenen Vorurteile zeigen (zumindest auf impliziten Maßen), obwohl sie alle sozial-kognitiven Fähigkeiten entwickelt haben. Weiterhin kann Aboud zwar einzelne Hinweise für ihre postulierten Entwicklungssequenzen und die damit verbundenen Entwicklung gruppenbezogener sozialer Einstellungen liefern, allerdings bleibt sie bis auf eine Ausnahme (Doyle & Aboud, 1995) längsschnittliche Untersuchungen, welche ihre Entwicklungssequenzen nachzeichnen, schuldig. Somit bleiben die von ihr vorgeschlagenen Entwicklungssequenzen auf dem Status weitestgehend unbestätigter Hypothesen. Schließlich vermuten einige Kritiker der SCT, dass die Reduktion der Vorurteile mit dem Alter eher einen Effekt der sozialen Erwünschtheit beziehungsweise der

Entwicklung intern repräsentierter sozialer Normen darstellt und nicht einer Entwicklung sozial-kognitiver Fähigkeiten. Diese Konkurrenzhypothese wurde allerdings bisher nicht geprüft.

Allerdings bietet Abouds Theorie günstige und einfache Ideen, um Interventionen zur Förderung von Intergruppenbeziehungen durchführen zu können. Die Förderung bestimmter von ihr identifizierter sozial-kognitiver Fähigkeiten, wie Perspektivenübernahmefähigkeit, Fähigkeit zur Akzeptanz fremder Meinungen, oder die multiple Klassifikationsfähigkeit stellen mögliche Interventionsansätze dar.

Social Identity Development Theory

Drew Nesdale (1999, 2000, 2004, 2008) betrachtet die Entwicklung von Vorurteilen nicht vor dem Hintergrund der kognitiven Entwicklung, wie es Frances Aboud tut. Seine Theorie basiert auf sozialpsychologischen Theorien, wie der Social Identity Theory (Tajfel und Turner, 1979) und der Self-Categorization Theory (Turner Hogg, Oakes, Reicher & Wetherell, 1987), sowie motivationalen Prozessen, wie sozialen Vergleichsprozessen in der Identitätsentwicklung. Seine zentrale These baut auf den oben genannten Theorien zur Entwicklung des sozialen Selbst auf. Vorurteile entwickeln sich bei Kindern, die sich auf Grund ihres Bedürfnisses nach einem positiven Selbstwert mit statushohen vorurteilsbelasteten Gruppen identifizieren. Mit dieser Identifikation übernehmen die Kinder die Vorurteile, welche in der jeweiligen Gruppe vorherrschen, um ihre positive Distinktheit gegenüber anderen Gruppen zu wahren.

Nach Drew Nesdales Social Identity Developmental Theory (SIDT) durchläuft diese Entwicklung von Vorurteilen bei Kindern vier Phasen. In der ersten undifferenzierten Phase (Alter: zwei bis drei Jahre) sind ethnische Merkmale typischerweise für Kinder nicht salient.

In der zweiten Phase erlangen die Kinder ab einem Alter von drei Jahren ein sogenanntes ethnisches Bewusstsein. Kinder können ab diesem Alter verschiedene Hautfarbtöne erkennen und differenzieren. Diese Wahrnehmung ergänzt um ein verbales Label, dass von Erwachsenen übernommen wird, ermöglicht die Kategorienbildung auf Grund von ethnischen Merkmalen. Kinder suchen sich dazu Kategorien, die in ihrer Gesellschaft eine soziale Bedeutung und Funktion haben. Die Entwicklung des ethnischen Bewusstseins ist ein Prozess der bis ins Alter von zehn bis elf Jahren andauern kann. Dabei ist eine erste wichtige Entwicklungsstufe die ethnische Selbstidentifikation. Dies lernen nach Nesdale alle Kinder der Majoritätsgruppe bis spätestens zum Alter von sechs oder sieben Jahren.

In der dritten Phase der Einstellungsentwicklung zeigen alle Kinder eine gewisse ethnische Präferenz. Sobald die ethnische Selbstidentifikation erreicht ist, und das Kind gelernt hat, dass es Mitglied einer ethnischen Kategorie ist, legt das Kind seinen Aufmerksamkeitsfokus sehr stark auf die Eigengruppe, achtet verstärkt auf Gemeinsamkeiten mit anderen Eigengruppenmitgliedern, hält die eigene Gruppe für relativ überlegen gegenüber anderen Gruppen und entwickelt eine Präferenz für die eigene Gruppe. Während Aboud (1988) sowie Tajfel und Turner (1979) davon

ausgehen, dass allein die Selbstkategorisierung ausreichend sei, um sowohl Eigengruppenpräferenz als auch Fremdgruppenabwertung zu entwickeln, sagt die SIDT, dass die Kategorisierung allein nur zur Präferenz für die Eigengruppe führt und nicht automatisch auch zu Abwertungen gegenüber der Fremdgruppe. Wobei Nesdale unter Vorurteilen einzig die negative Bewertung und Diskriminierung von Mitgliedern anderer Gruppen versteht.

In der vierten und letzten Phase zeigen die Kinder frühestens ab einem Alter von sieben Jahren schließlich ethnische Vorurteile. Gerade in diesem Zeitraum kommt es nach der SIDT und konträr zur SCT Abouds zur Entwicklung von Vorurteilen; allerdings nur bei den Kindern, die überhaupt Vorurteile entwickeln. Die Entwicklung von Vorurteilen gegenüber der Fremdgruppe erfordert dazu eine Verschiebung des Aufmerksamkeitsfokus von der Eigengruppe auf die Fremdgruppe, zumindest aber eine Aufteilung der Aufmerksamkeit zu gleichen Teilen. Drei Faktoren beeinflussen den Übergang von Phase drei „Ethnische Präferenz“ zu Phase vier „Ethnische Vorurteile“:

Zunächst müssen die Kinder die Stufe der „Ethnischen Konstanz“ erreicht haben. Ethnie wird dann als stabiles, nicht veränderbares Merkmal erkannt, und kann somit als Anker für die Anlagerung von Eigenschaftsinformationen für ethnische Gruppen dienen. Weiterhin spielen einige sozial-kognitive Fähigkeiten, wie Perspektivenübernahme, Empathie, oder die Moralentwicklung eine wichtige Rolle. Schließlich sind nach Nesdale besonders soziale Identitätsprozesse zentral für die Ausprägung von Vorurteilen bei Kindern. Die negativen Einstellungen der Eigengruppe gegenüber der Fremdgruppe werden dabei vom Kind übernommen, wenn diese Gruppe zum eigenen Selbstbild des Kindes passt und eine Möglichkeit zu positiver Distinktheit und somit zu einem positiven Selbstwert des Kindes bietet. Dieser Vorgang wird wiederum durch drei Faktoren befördert: 1. Die Identifikation mit einer Gruppe, die Vorurteile pflegt. 2. Eine Spannung oder Bedrohungssituation zwischen der Eigengruppe und einer Fremdgruppe. 3. Ein Konflikt zwischen der Eigengruppe und einer Fremdgruppe.

Die Theorie von Nesdale birgt einige zentrale Implikationen. So sind nach der SIDT die Einstellungen der Kinder nicht zwingend abhängig von den Einstellungen der Eltern, sondern spiegeln eher die Interessen und Wahrnehmungen der Kinder selbst wider. Weiterhin folgt aus der Theorie, dass soziale und motivationale Faktoren den Einfluss von sozial-kognitiven Variablen, wie z.B. der Fähigkeit zur

Perspektivenübernahme begrenzen. Als dritte Implikation folgt aus der SIDT, dass Kinder ab einem Alter von etwa sechs Jahren im weiteren Entwicklungsverlauf immer Vorurteile entwickeln können. Dies ist einzig abhängig von der sozialen Situation und der vorherrschenden Gruppe mit der sich das Kind identifiziert. Nicht zuletzt sieht Nesdale es nicht als zwingend an, dass ein Kind oder ein Erwachsener überhaupt Vorurteile entwickelt, solange es bzw. er sich nicht mit einer vorurteilsbehafteten Gruppe identifiziert.

Empirisch lässt sich das Modell von Nesdale durchaus stützen. So konnten Bigler und Kollegen zeigen, dass Kinder selbst in einem willkürlich zugeordneten minimalen Gruppenparadigma bereits nach wenigen Wochen eine Eigengruppenpräferenz entwickeln (Bigler, 1995; Bigler, Jones & Lobliner, 1997). Auch der soziale Vergleich und der damit verbundene Status einer Gruppe spielen für Kinder bereits eine Rolle. Nesdale und Flesser (2001) konnten zeigen, dass bereits Fünfjährige verstärkte Eigengruppenpräferenz zeigen, wenn sie einer Gruppe mit hohem Status angehören bzw. den Wunsch haben ihre Gruppe zu wechseln, falls diese einen niedrigen Status aufweist.

Kinder ab einem Alter von fünf Jahren sind sich also Statusunterschieden zwischen Gruppen bewusst und möchten gerne der Gruppe mit hohem Status angehören. Weiterhin zeigen die oben genannten Studien, dass allein durch die Selbstkategorisierung die Kinder Mitglieder ihrer eigenen Gruppe mehr mögen als die Mitglieder fremder Gruppen. Diese werden allerdings nur weniger gemocht und nicht abgewertet, was der Phase der ethnischen Präferenz Nesdales entspricht.

Es scheint also die saliente Selbst- und Fremdkategorisierung zu sein, welche den Effekt der Eigengruppenpräferenz bei Kindern auslöst. Diese Kategorisierung kann unabhängig von der Ethnie der Kinder sein und trotzdem zu einer Einstellungsverzerrung führen (Nesdale, Maas, Griffiths & Durkin, 2003).

Die Stärke der SIDT liegt insgesamt in der Berücksichtigung der sozialen und motivationalen Prozesse bei der Entwicklung von Vorurteilen. Dies sind unter anderem soziale Vergleichsprozesse, Statusunterschiede zwischen Gruppen und Auswahl von sowie Identifikation mit Gruppen zur Steigerung und Stabilisierung des kindlichen Selbstwerts. Allerdings beschränkt sich die bisherige Forschung hauptsächlich auf die Phase der ethnischen Präferenz und belässt die anderen Phasen weitgehend unbeleuchtet.

Da nach der Social Identity Developmental Theory Kinder nur dann Vorurteile entwickeln, wenn sie sich mit Gruppen identifizieren, die ebenfalls Vorurteile als Gruppennorm haben, müssten Interventionen derart gestaltet sein, dass sie Kinder dazu anregen, sich mit Gruppen ohne Vorurteile zu identifizieren. Weiterhin sollten Vorurteile und Diskriminierung in der Gesellschaft geächtet werden, um somit nur erschwert als Gruppennorm etabliert werden zu können.

Ein anderer Ansatzpunkt, der sich aus dieser Theorie ableiten lässt, sind die Kategorisierungsprozesse, welche einer Identifikation mit einer Gruppe vorausgehen. Hier könnte über Dekategorisierungs- und Rekategorisierungsprozesse versucht werden, eine individualisierte Personenwahrnehmung zu etablieren (s.u. Kapitel II.3.2. Kontakt-Interventionen und Kapitel II.3.4. Wissensbasierte Interventionen).

Integrative Developmental Contextual Theory

In den letzten zwanzig Jahren wurde die Entwicklung und der Entwicklungsverlauf von Vorurteilen bei Kindern hauptsächlich mit den zwei oben erläuterten Theorien erklärt: Der SCT (Aboud, 1988) und der SIDT (Nesdale, 1999, 2000). Beide Theorien können Teile der Entwicklung von Vorurteilen und Stereotypen erklären und ihre Vorhersagen mit empirischen Studien stützen. Allerdings können beide Theorien nicht alle Ergebnisse empirischer Studien zur Vorurteilsentwicklung erklären. So ließ sich beispielsweise in israelischen Studien, im Kontext eines dauerhaften Konfliktes, kein Rückgang von Vorurteilen bei Kindern im Alter zwischen acht und neun Jahren feststellen, wie es die SCT voraussagt und auch in friedlichen multiethnischen Kontexten belegen konnte (Bar-Tal und Teichman, 2005). Diese und andere bisher nicht erklärbare Ergebnisse veranlassten Daniel Bar-Tal und Yona Teichman eine eigene Theorie zur Entwicklung von Vorurteilen und Stereotypen zu entwickeln. Diese Theorie soll insbesondere den vorliegenden sozialen Kontext, wie z.B. einen dauerhaften gewalttätigen Konflikt zwischen Gruppen, in das Modell der Vorurteilsentwicklung mit einbeziehen.

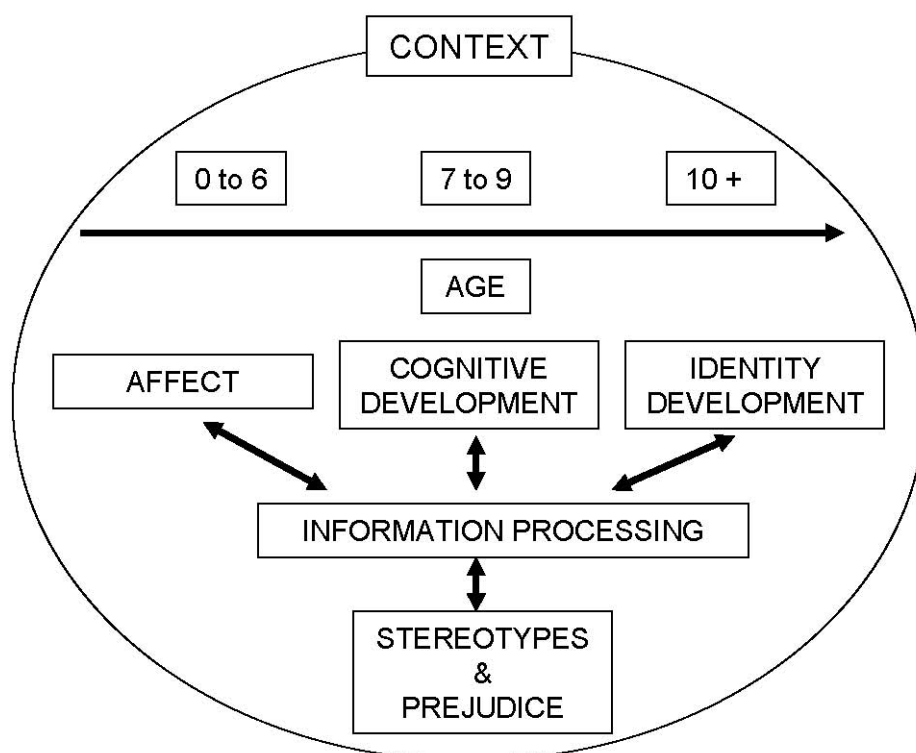
In der Integrative Developmental-Contextual Theory (IDCT) wird davon ausgegangen dass die Entwicklung der sogenannten SPIRs² (Shared Psychological Intergroup Repertoire), also der gruppenbezogenen sozialen Einstellungen, durch das Zusammenspiel mehrerer Faktoren bedingt wird: Zum einen von sozial-kognitiven Prozessen (wie in der SCT) zum anderen von selbstwerterhöhenden Prozessen (wie in der SIDT) und zum dritten vom Affekt der Kinder. Diese drei Faktoren betten die Autoren in die Theorie zur Identitätsentwicklung von Erikson (Psychosoziales Entwicklungsmodell, 1968) und Marcia (1980, 1998) ein. Alle drei Faktoren haben demnach eine kritische Phase während der Entwicklung, in denen sie besonders bedeutsam und aktiv sind. Allerdings wirken sich alle drei Faktoren

² Teichman und Bar-Tal gehen davon aus, dass gesellschaftliche Intergruppen-Konflikte eine sozio-psychologische Infrastruktur erzeugen, die Teil der Konfliktkultur wird. Diese beinhaltet Erzählungen, Überzeugungen, Einstellungen und Emotionen welche mit der Ursache des Konflikts sowie mit dessen Verlauf, Zielen und eventuellen Lösungen verknüpft werden. Die Infrastruktur dient als eine Art Prisma durch das die Beteiligten den Konflikt, sich selbst und die Gegenpartei wahrnehmen, um den Konflikt im Alltag besser bewältigen zu können. Diese sozio-psychologische Infrastruktur nennen Teichman und Bar-Tal SPIR: Shared Psychological Intergroup Repertoire.

bereits von der jüngsten Kindheit an und über die gesamte Entwicklungsspanne hin aus. Dabei werden die Faktoren von spezifischen Kontexteinflüssen wie z.B. einem dauerhaften Konflikt genährt, also gefördert oder gehemmt. Einmal entwickelt wirken die SPIRs sich wiederum auf die oben genannten Entwicklungsfaktoren und auf die soziale Informationsverarbeitung der Kinder aus. Ein Modell zur Integrative Developmental-Contextual Theory findet sich unten (Abbildung 01).

Abbildung 01

Modell der Integrative Developmental-Contextual Theory (Teichman & Bar-Tal, 2008)



Im Folgenden werden die nach der IDCT zentralen Entwicklungsfaktoren, die die Entwicklung von Vorurteilen beeinflussen beschrieben.

Affekt

Aboud (1988) beschreibt den durch Personen ausgelösten Affekt als entscheidend für basale Reaktionen auf diese Personen (positiv wie negativ). Sie geht davon aus, dass diese affektiv gesteuerten Reaktionen im Alter von drei bis sechs Jahren das Verhalten des Kindes dominieren und danach geringer werden. Die IDCT geht aber davon aus, dass der ausgelöste Affekt auch im späteren

Entwicklungsverlauf bedeutsamen Einfluss auf Vorurteile und Stereotype behält, indem beispielsweise durch negative Gefühle gegenüber einer Person oder einer Gruppe, der Prozess der Informationsverarbeitung beeinflusst und verzerrt wird. So können neue Informationen „übersehen“ werden und Fremdgruppenmitglieder weiterhin anhand von Stereotypen und Vorurteilen beurteilt werden. Dies führt, unabhängig von der Entwicklung in anderen Bereichen (Kognitionen, Identität) zur Stabilisierung und Aufrechterhaltung derartiger Einstellungen (Holt & Silverstein, 1989).

Somit spielen Emotionen in der frühen Kindheit aber auch über die gesamte Lebensspanne hinweg eine wichtige Rolle bei dem Erwerb, der Aufrechterhaltung und der Veränderung von interpersonalen und intergruppalen Einstellungen und Verhaltensweisen.

Kognitive Entwicklung

Die kognitive Entwicklung findet sich häufig als Korrelat der Entwicklung von Einstellungen. So kann ein normativer Entwicklungsverlauf der Einstellungsentwicklung mittels der kognitiven Entwicklung erklärt werden (s.o. SCT). Allerdings können manche empirische Ergebnisse nicht von der SCT erklärt werden. So fanden Bar-Tal und Teichman (2005) in ihren Stichproben die ersten Intergruppeneinstellungen bereits sehr früh in der Kindheit. Zudem konnten sie keine Abnahme der Vorurteile mit der Entwicklung des konkret operationalen Denkens der Kinder feststellen. Sie vermuten, dass in ihrem Fall der Einfluss des Kontextes des dauerhaften Konfliktes zwischen Juden und Arabern den Einfluss der kognitiven Entwicklung überwindet. Sie sehen in der kognitiven Entwicklung auch einen wichtigen Faktor bei der Entwicklung von Vorurteilen. Dieser sei allerdings abhängig vom Kontext in dem die Kinder aufwachsen sowie von den anderen beteiligten Faktoren.

Identitätsentwicklung

Während der prä- und frühen Adoleszenz wird die Identitätsentwicklung die wichtigste Entwicklungsaufgabe der Kinder. Ein starkes Bedürfnis nach Selbstwerterhöhung spielt in dieser unsicheren Findungsphase eine besondere Rolle für die Intergruppeneinstellungen. Dieses Bedürfnis wird über die Identifikation mit einer statushohen Eigengruppe befriedigt (s.o. SIDT). Im Kontext eines dauerhaften

Konflikts tritt dieses Bedürfnis besonders auf und führt zu einer Zunahme der Eigengruppenpräferenz sowie der Fremdgruppenabwertung (Bar-Tal & Teichman, 2005).

Diese drei Entwicklungsfaktoren (Affekt, kognitive Entwicklung, Identitätsentwicklung) wirken jedoch nicht im sozialen Vakuum, sondern im Wechselspiel mit einem spezifischen sozialen Kontext. Was die Autoren unter dem Begriff sozialer Kontext verstehen und wie dieser sich auf die Vorurteilsentwicklung auswirken kann, wird im Folgenden kurz erläutert.

Einfluss des sozialen Kontext auf SPIRs

Ashmore, Deaux und McLaughlin (2004) definieren sozialen Kontext als „general and continuing multilayered and interwoven set of material realities, social structures, and shared belief system that surround any situation” (pp. 103). Für die Intergruppenperspektive ist besonders die Ebene der „shared beliefs”, der geteilten Überzeugungen relevant. Diese sind Teil des kulturellen Kontextes und der SPIRs. Die SPIRs werden durch Sozialisationsagenten (Eltern, Medien) an die nächste Generation weitergegeben (Bourdieu, 1977). Die meisten Erwachsenen teilen die gleichen SPIRs (Oppenheimer, 2006) und diese werden mit der Zeit rigide und resistent gegen Veränderungen. Obwohl die SPIRs und der kulturelle Kontext stabil sind, können sie durch bedeutende Ereignisse verändert werden. Solche Ereignisse sind beispielsweise Kriege oder dauerhafte Konflikte.

Nach ihrer Theorie kommen Bar-Tal und Teichman zu zwei unterschiedlichen kontextabhängigen Vorhersagen zu der Entwicklung von Vorurteilen. In neutralen multiethnischen Kontexten sei der Entwicklungsverlauf ähnlich wie ihn Frances Aboud und Drew Nesdale beschreiben. In Kontexten mit dauerhaften Konflikten allerdings sei dieser Entwicklungsverlauf intensiviert oder verändert, da die Kinder von klein auf eine Bedrohung durch eine spezifische fremde Gruppe spüren. Dies führt zu einer früheren Entwicklung der entsprechenden sozialen Kategorien sowie einer stark verzerrten Wahrnehmung der Eigen- und Fremdgruppe. In diesen Fällen läuft die gleiche Entwicklungssequenz entweder verdichtet und intensiver ab, oder der Einfluss des Kontextes ist so stark dass die normalen kognitiven und altersbedingten Einflüsse durch ihn überdeckt werden und es zu einem anderen Entwicklungsverlauf kommt.

Dies zeigt sich auch in einigen empirischen Studien. So entwickeln jüdisch-israelische Kinder schon früher Vorformen gruppenbezogener sozialer Einstellungen, wie negativen Affekt (Gefühle der Angst oder Bedrohung), bei einer Konfrontation mit Mitgliedern der arabischen Fremdgruppe. Dieses Gefühl der Bedrohung zieht sich bis ins junge Erwachsenenalter hin (16-17 jährige) (Teichman & Bar-Tal, 2008). Prozesse wie soziale Kategorisierungen und Differenzierungen treten bei jüdischen Kindern in Israel ebenfalls früher auf als in anderen konfliktfreien Forschungskontexten (Teichman & Bar-Tal, 2008). Auch der Rückgang von gruppenbezogenen sozialen Einstellungen in der mittleren bis späten Kindheit konnte im israelischen Kontext nicht bestätigt werden. Hier wurde eher ein weiterer Anstieg der Eigengruppenpräferenz und Fremdgruppenarbwertung vorgefunden (Bar-Tal und Teichman, 2005, Assaf, 2006). Allerdings scheint dies vor allem dann der Fall zu sein, wenn eine Eskalation des Konflikts vorliegt, somit scheint die Entwicklung abhängig vom Status des herrschenden Konflikts zu sein (Assaf 2006, Teichman 2001, Teichman & Bar-Tal, 2008).

Der Vorteil der IDCT ist die multifaktorielle Betrachtung der Entwicklung von Vorurteilen: Emotionale Erfahrungen, kognitive Entwicklung und die Identitätsentwicklung werden im Zusammenspiel mit dem vorherrschenden sozialen Kontext betrachtet. Dies ergibt keinen allgemeingültigen Entwicklungsverlauf, sondern immer eine kontextspezifische Entwicklung. Weiterhin haben die Autoren die Entwicklung von Vorurteilen vom Kleinkindalter bis hin zur Adoleszenz untersucht, was ebenfalls eine Stärke dieser Theorie darstellt. Allerdings basieren die Ergebnisse ausschließlich auf Studien für den Kontext Israel mit dem Konflikt zwischen Juden und Arabern, was die Generalisierbarkeit (auch auf andere Konfliktregionen wie Nordirland) der Theorie und ihrer Ergebnisse noch offen lässt.

Für Interventionen bietet die IDCT ähnliche Ansätze wie die Theorien von Aboud und Nesdale. Dabei sollte jedoch auf zwei Dinge besonders geachtet werden: Zum einen auf den sozialen Kontext in dem die Kinder leben und in dem die Intervention stattfindet (besonders auf die vorherrschenden Intergruppenbeziehungen), zum anderen auf das jeweilige Entwicklungsstadium und damit auf die Sensitivität der teilnehmenden Kinder für bestimmte Prozesse (Affekte, Kognitionen, Identität).

Jüngere Kinder sollten verstärkt über affektive Erfahrungen erreicht werden. Kinder im mittleren Alter können besonders von kognitiven Übungen profitieren und ältere Kinder sollten in Bezug auf ihre Identitätsentwicklung und ihren Selbstwert hin gefördert werden.

Zusammenfassung der aktuellen Theorien

Die hier vorgestellten Theorien zur Entwicklung von Vorurteilen und Stereotypen bei Kindern sollen an dieser Stelle nochmals kurz zusammengefasst und kontrastiert werden:

Frances Aboud kann mit ihrer Social Cognitive Developmental Theory of Prejudice sicherlich als Vorreiterin der „modernen“ Erforschung der Vorurteilsentwicklung betrachtet werden. Sie hat die Idee, dass die Entwicklung von Vorurteilen mit der sozial-kognitiven Entwicklung des Kindes verknüpft ist und baut dabei auf dem Modell der kognitiven Entwicklung Jean Piagets auf. Defizite in der sozial-kognitiven Entwicklung, wie z.B. in den Fähigkeiten zur Perspektivenübernahme (Aboud, 1988), zur multiplen Klassifikation (Bigler & Liben, 1993), oder zur differenzierten Betrachtung von Fremdgruppenmitgliedern (Aboud & Doyle, 1996b) gehen mit verstärkten Vorurteilen gegenüber Fremdgruppen einher. Kognitive Limitationen im Kindesalter stützen also die Entwicklung von Vorurteilen. Aboud ist somit eine entwicklungspsychologische Perspektive auf das Thema zu verdanken.

Im Gegensatz dazu vertritt Drew Nesdale eine eher motivational-sozialpsychologische Richtung bei der Thematisierung der Vorurteilsentwicklung. Seine Social Identity Developmental Theory basiert auf der Social Identity Theory von Tajfel und Turner (1979) sowie auf der Self-Categorization Theory (Turner et al., 1987). Er differenziert klar zwischen Eigengruppenpräferenz und Fremdgruppenabwertung, wobei nur Letzteres für ihn als „echtes“ Vorurteil gilt. Die Eigengruppenpräferenz beschreibt er als Teil einer normalen kindlichen Entwicklung, während die Fremdgruppenabwertung, also die eigentlichen Vorurteile, nur unter bestimmten sozialen Bedingungen, wie der Identifikation des Kindes mit einer vorurteilsbelasteten Gruppe, auftritt. Faktoren wie soziale Vergleichsprozesse, Gruppenstatus, und die Motivation den eigenen Selbstwert zu erhalten oder zu erhöhen liegen nach Nesdale der Vorurteilsentwicklung zu Grunde.

Daniel Bar-Tal und Yona Teichman erkennen den Nutzen der beiden erstgenannten Ansätze. Sie integrieren die Social Cognitive Developmental Theory of Prejudice von Aboud und die Social Identity Developmental Theory Nesdales.

Weiterhin weisen sie auf die Bedeutung des Affekts hin, um Einstellungsentwicklungen zu erklären. Diese Einflussfaktoren (Affekt, kognitive Entwicklung, soziale Identitätsentwicklung) integrieren die Autoren in das Modell zur Identitätsentwicklung von Erikson (Psychosoziales Entwicklungsmodell, 1968), beziehungsweise Marcia (1980, 1998). Sie betrachten die Entwicklung von Vorurteilen bei Kindern somit als einen multifaktoriellen Prozess mit sequentiell verschiedenen Schwerpunkten. Dabei heben die Autoren hervor, dass dieser Prozess jedoch immer unter dem Einfluss des vorliegenden sozialen Kontextes (wie z.B. einem dauerhaften Konflikt zwischen zwei Gruppen) zu betrachten ist, da der Kontext und die Entwicklungsfaktoren sich wechselseitig bei der Entwicklung von Vorurteilen beeinflussen. Damit bieten Bar-Tal und Teichman einen umfassenden Ansatz zur Erforschung der Entwicklung von Vorurteilen bei Kindern.

Schlussfolgerungen aus der Entwicklung von gruppenbezogenen sozialen Einstellungen

Insgesamt lässt sich zur Entwicklung von Vorurteilen festhalten, dass Konsens darüber besteht, dass die ersten Grundsteine solcher Einstellungen bereits sehr früh in der Kindheit im Alter von zwei bis drei Jahren gelegt werden. Weiterhin scheint die Phase der mittleren bis späten Kindheit (fünf bis zwölf Jahre) besondere Bedeutung für die Einstellungsentwicklung zu haben. Hier spielen besonders die kognitive Entwicklung aber auch Faktoren der Identitätsentwicklung sowie kontextbezogene Faktoren eine wichtige Rolle.

Die dargestellten Theorien bergen für die Entwicklung von Interventionen zur Prävention oder Reduzierung von Vorurteilen nützliche Richtlinien. So sollten Interventionsprogramme in der frühen Kindheit, etwa bis zum sechsten Lebensjahr, besonders die emotionalen Aspekte von gruppenbezogenen sozialen Einstellungen ansprechen. Ab dem sechsten Lebensjahr sollte auf Grund der fortschreitenden kognitiven Entwicklung der Kinder mehr auf die sozial-kognitiven Faktoren, wie zum Beispiel die Förderung der Fähigkeit zur Perspektivenübernahme, in Interventionsprogrammen eingegangen werden. Schließlich sollten Interventionsprogramme ab dem zehnten Lebensjahr verstärkt auf Bereiche der sozialen sowie persönlichen Identitätsentwicklung abzielen. Zusätzlich sollte jedes Interventionsprogramm den sozialen Kontext der Zielgruppe, die erreicht werden soll, berücksichtigen.

II.3. Empirische Studien

II.3.1. Klassifikation von Interventionen

Es gibt verschiedenste Ansätze und Ideen um Vorurteile und Diskriminierung zu reduzieren. Genauso gibt es unterschiedliche Möglichkeiten diese Ansätze zu kategorisieren. Im Folgenden werden einige Einteilungen von verschiedenen Autoren vorgestellt und das für diese Arbeit sinnvollste Kategoriensystem eingeführt.

Interventionsansätze zur Reduktion von Vorurteilen lassen sich beispielsweise nach Theorien einteilen, die den entsprechenden Ansätzen zu Grunde liegen. Pfeifer und Kollegen (Pfeifer, Spears Brown & Juvonen, 1999) teilen in ihrem Überblicksartikel Interventionen nach der Sozialisationstheorie (a), der Kontakthypothese (b) und der Social Cognitive Developmental Theory of Prejudice (c) ein. Dabei nennen sie als Beispiele für entsprechende Interventionen a) multikulturelle Curricula (z.B. Litcher & Johnson, 1969; Singh & Yancey, 1974; Katz & Zalk, 1978), b) Kooperative Lerntechniken (z.B. Ziegler, 1981; Johnson & Johnson, 1981) und c) die Förderung sozial-kognitiver Kompetenzen wie der Fähigkeit zur multiplen Klassifikation (z.B. Bigler & Liben, 1992) oder von Perspektivenübernahme und Empathie (z.B. Gimmestad & DeChiara, 1982).

Eine etwas differenziertere theorieorientierte Unterteilung wählen Aboud und Levy (2000). Sie unterscheiden zwischen integrativen Schulen, bilingualer Erziehung (basieren beide auf Kontakthypothese), multikulturellen Curricula (basiert auf Sozialisationstheorie), sowie sozial-kognitiven Trainings und Rollenspielen bzw. Empathie-Übungen (welche beide auf der sozial-kognitiven Entwicklungstheorie basieren).

Andere Einteilungen beziehen sich auf die Ursachen von Vorurteilen und die daraus resultierenden Ebenen auf denen die Interventionen ansetzen. So beschreibt Duckitt (1992a) vier Ebenen als Ansatzmöglichkeiten für Interventionen gegen Vorurteile: 1. Genetische und evolutionäre Prädisposition. 2. Gesellschaftliche, organisationale und intergrupale Muster für Kontakt und Normen zu Intergruppenbeziehungen – also Gesetze oder Normen zu Segregation, Machtverteilung zwischen Gruppen und Intergruppenverhalten etc.. 3. Mechanismen des sozialen Einflusses auf Gruppenebene oder interpersonaler Ebene – z.B.

Einflüsse der Medien, des Bildungssystems oder der Struktur von Organisationen. 4. Interindividuelle Unterschiede in der Anfälligkeit für Vorurteile und diskriminierendes Verhalten – wie Persönlichkeitseigenschaften, kognitive Stile oder grundlegende Schemata. Dazu stellt Duckitt fest:

„The higher the level of intervention [i.e., lower numbers], the greater will be its potential impact” (S.251).

Eine weitere Möglichkeit Interventionen einzuteilen, besteht in der Fokussierung auf die Prozessebene. Das bedeutet die Mechanismen, welche zur Veränderung der Einstellungen anvisiert werden, erhalten eine Unterteilung, wie z.B. Oskamp (2000) Interventionen nach der kognitiven Ebene, der behavioralen Ebene und der motivationalen Ebene unterteilt.

Weiterhin kann man zwischen normativen und informativen Programmen unterscheiden (Deutsch & Gerard, 1995), wobei normative Programme versuchen, positive soziale Normen für den Umgang mit anderen Gruppen zu installieren, während informative Interventionen kognitives Wissen vermitteln.

Auch der Grad an Aktivität der Teilnehmer kann zur Differenzierung von Interventionsprogrammen genutzt werden. Dabei scheinen aktive bzw. interaktive Programme stärkere und länger andauernde Effekte auf Einstellungen und Verhalten zu zeigen als passive Programme, wie das reine Aufnehmen von Informationen (Oskamp, 1991).

Eine weitere Klassifikation interventiver Maßnahmen zur Reduktion oder Prävention von Vorurteilen schlagen Beelmann und Kollegen vor (Beelmann, Heinemann & Saur, 2009). Sie unterscheiden die vorliegenden Interventionsansätze nach zwei Dimensionen: Den Interventionsinhalten und den Interventionsstrategien.

Auf der inhaltlichen Dimension werden alle Interventionen auf den drei Kategorien Kontakt-Interventionen, wissensbasierte Intergruppenintervention und individuelle Kompetenzförderung abgebildet. Die Dimension der Interventionsstrategien unterteilen die Autoren in Bildungs- und Erziehungsmaßnahmen, standardisierte Programme und mediengestützte Maßnahmen. Somit ergibt sich folgende Matrix an Interventionsansätzen (s. Tabelle 02).

Tabelle 02

Klassifikation von Interventionen zur Prävention negativer gruppenbezogener sozialer Einstellungen

Interventionsinhalte	Interventionsstrategien		
	Bildungs- und Erziehungsmaßnahmen	Standardisierte Programme	Mediengestützte Maßnahmen
Kontaktinterventionen	Integratives Schulwesen, Kooperatives Lernen	Stellvertretender-Kontakt-Programme, Koexistenzprogramme	Medienkampagnen mit prominenten Identifikationsfiguren
Wissensbasierte Intergruppeninterventionen	Politische Bildungsmaßnahmen	Multikulturelle Trainingsprogramme, Antirassismusprogramme	Informationen im Rahmen von Kulturangeboten
Individuelle Kompetenzförderung	Bildungsauftrag in der pädagogischen Standardversorgung	Kognitives Training, Empathietraining, Zivilcouragetraining	Filme, Bücher und andere Materialien mit pädagogischem Anspruch

Anmerkungen. Aus Beelmann, Heinemann & Saur, 2009 (S. 437)

Dieses Klassifikationssystem ist für die Einordnung von PARTS besonders geeignet. Das in dieser Arbeit zu evaluierende Programm PARTS ist multimodal und setzt sich aus Elementen aus allen drei von Beelmann und Kollegen vorgeschlagenen Interventionsinhalten zusammen. Dabei kann PARTS interventionsstrategisch als standardisiertes Programm eingeordnet werden.

Es ist daher festzustellen, dass es verschiedenste Ansätze und Dimensionen zur Kategorisierung von Interventionen zur Reduktion von Vorurteilen gibt. Für die vorliegende Arbeit ist eine inhaltliche Kategorisierung besonders sinnvoll, da es sich um die Evaluation eines standardisierten multimodalen (also multi-inhaltlichen) Programms handelt. Die drei zentralen Inhaltsbereiche der Intervention sind dabei: Kontakt-Interventionen, wissensbasierte Intergruppeninterventionen und die Förderung individueller Kompetenzen.

Im Folgenden werden die zentralen empirischen Studien aus diesen drei inhaltlichen Interventionsbereichen zur Prävention von negativen gruppenbezogenen sozialen Einstellungen ausführlich vorgestellt. Dazu wird zunächst auf die der jeweiligen Methode zu Grunde liegenden Wirkmechanismen eingegangen. Im Anschluss daran erfolgt ein Bericht der zentralen Studien.

II.3.2. Kontakt-Interventionen

Beispiele für Kontakt-Interventionen zur Förderung von Intergruppenbeziehungen finden sich zahlreich. Angefangen von desegregierten Schulen oder Nachbarschaften, über bilinguale Beschulungsprogramme bis hin zu kooperativen Lerntechniken besteht eine große Bandbreite möglicher Kontakt-Interventionen.

Sie alle beruhen auf einer zentralen These bei der Erforschung und Reduzierung von Vorurteilen und Diskriminierung: Auf der Kontakt-Hypothese. Im Kern beinhaltet diese Hypothese, dass Kontakt zwischen Gruppen zu einer Reduzierung von Vorurteilen und einer Verbesserung der Intergruppenbeziehungen führt. Besonders Allport (1954) hat mit seinem Buch „The Nature of Prejudice“ einen zentralen Beitrag zur Kontakt-Hypothese geliefert. Er beschreibt vier Faktoren, welche eine Kontaktsituation zwischen Gruppen in ihrer Wirksamkeit gegen Vorurteile maßgeblich beeinflussen: Gleicher Status zwischen den Gruppen; gemeinsame Ziele; Kooperation zwischen den Gruppen und der unterstützende Einfluss von Autoritäten, Normen, Gesetzen und Bräuchen.

Diese Forschungsrichtung hat in der Sozialpsychologie eine lange Tradition und im Laufe der Zeit ist durch zahlreiche empirische Untersuchungen eine umfangreiche Datenmenge zu diesem Thema entstanden. Thomas Pettigrew und Linda Tropp (2006) widmeten sich in ihrem Artikel der systematischen Auswertung und Zusammenfassung all dieser bisher gesammelten und veröffentlichten Daten. Sie wollen mit ihrer Metaanalyse die Auswirkungen von Intergruppenkontakt auf Vorurteile prüfen. Dazu ließen sie die Daten von 713 unabhängigen Stichproben aus 515 Studien in ihre Analyse einfließen. Intergruppenkontakt ist in diesen Studien als direkter face-to-face Kontakt zwischen Mitgliedern diskreter Gruppen definiert. Die in die Analyse eingehenden Studien beziehen sich dabei nicht nur auf den Kontakt und die Vorurteile zwischen ethnischen Gruppen, sondern auch auf andere soziale Gruppen wie Senioren, behinderte Menschen oder psychisch Kranke. Generell finden die Autoren einen positiven Effekt von Intergruppenkontakt auf Vorurteile. Die Effektstärken bewegen sich dabei abhängig von der Untersuchungseinheit (Studie, Stichprobe oder Tests) und von dem Berechnungsmodell (fixed vs. random) in einem Bereich von $r = -.21$ bis $r = -.23$. Dieser Effekt bleibt stabil auch wenn verschiedene Gefährdungen der Validität, wie die kausale Interpretierbarkeit, der Publikationsbias, die Generalisierbarkeit und die Qualität der Studien, kontrolliert werden.

Weiterhin können Pettigrew und Tropp zeigen, dass die von Allport spezifizierten Kontaktbedingungen einen positiven Einfluss auf den Kontakteffekt haben. Dazu haben sie einen generellen „Programmfaktor“ in ihre Metaanalyse eingeführt, der Studien kennzeichnet, in welchen besonders auf die Kontaktbedingungen geachtet wurde. Dabei ergibt die Analyse der 134 Stichproben mit optimalen Kontaktbedingungen einen stärkeren Einfluss auf Vorurteile ($r = -.29$) als bei den restlichen 562 Stichproben mit unstrukturiertem Kontakt ($r = -.20$). Diesen Effekt der Kontaktbedingungen können die Autoren auch in einer Regressionsanalyse nachweisen, in der der Effekt anderer Qualitätsmerkmale der Studien, wie die Qualität der Messungen von abhängigen und unabhängigen Variablen, die Kontaktintensität der Kontrollgruppe, oder die Art der Studie, kontrolliert wird. Welche von Allports vier Bedingungen dabei eine besondere Rolle spielt, lässt sich auch bei differenzierteren Analysen allerdings nicht festmachen.

In dem Artikel von Tropp und Prenovost (2008) untersuchen die Autoren gezielt Kontakteffekte in Studien mit Kindern und Jugendlichen. Auch hier werden Ergebnisse aus dem Pool der Metaanalyse von Pettigrew und Tropp (2006) metaanalytisch zusammengefasst und ausgewertet. Insgesamt ergibt sich bei Kindern und Jugendlichen ebenfalls ein negativer Zusammenhang $r = -.22$ (Pettigrew und Tropp, 2006) zwischen Kontakt und Vorurteilen. Je nach Alter der Teilnehmer der Studien ist der Effekt allerdings unterschiedlich stark. Bei Kindern bis 12 Jahre ergibt sich ein negativer Zusammenhang von $r = -.24$. Bei Jugendlichen zwischen 13 und 17 Jahren sinken die Kontakteffekte etwas ab ($r = -.21$) und steigen bei Collegestudenten (18 bis 21 Jahre) wieder leicht an ($r = -.23$).

Dabei unterscheiden sich die Zusammenhänge für Kinder und Jugendliche nicht bedeutsam nach dem Kontext (Schule vs. zu Hause) oder der Zielgruppe (Ethnie, Behinderte, Senioren) der Studien. Der Kontakteffekt scheint also auch bei Kindern und Jugendlichen generell wirksam zu sein. Allports (1954) Bedingungen für optimalen Kontakt zwischen Gruppen machen aber auch hier einen Unterschied aus: Wenn die Bedingungen erfüllt sind steigt die Kontaktwirkung auf $r = -.29$. Wenn der Kontakt unstrukturiert verläuft, sind die Ergebnisse etwas schwächer ($r = -.20$). Speziell im Schulsetting wirkt sich die Einhaltung der Kontaktbedingungen positiv auf die Wirksamkeit des Kontakts aus ($r = -.34$ Bedingungen eingehalten vs. $r = -.19$ unstrukturierter Kontakt).

Insgesamt kann festgehalten werden, dass Kontakt zwischen Gruppen Vorurteile auch bei Kindern und Jugendlichen reduziert. Dies gilt in besonderem Maße wenn der Kontakt nach den Bedingungen für optimalen Kontakt nach Allport aufgebaut wird.

Wirkprozesse von Kontakt-Interventionen

Die Mechanismen, die für die Wirksamkeit von intergruppalen Kontakt auf die Intergruppeneinstellungen verantwortlich sind, sind bisher weniger gut erforscht. Dovidio und Kollegen (Dovidio, J.F., Gaertner, S.L., Stewart, T.L., Esses, V.M., Vergert, M.T., & Hodson, G., 2004) vermuten besonders affektive Prozesse, die mediierend zwischen Kontakt-Interventionen und Einstellungsveränderungen wirksam sind. Im Folgenden werden diese kurz vorgestellt:

a) Aufbau von positivem Affekt

So ermöglicht Kontakt, Empathie, oder eine affektive Verbindung allgemein, zu anderen aufzubauen. Dies führt zu positiveren Gefühlen diesen anderen gegenüber und verringert den intergruppalen Bias. Batson und Kollegen (Batson, Early & Salvarani, 1997) konnten dazu zeigen, dass Personen mit Fokus auf die Emotionen von anderen, diesen Personen und deren Gruppe positivere Gefühle entgegenbringen. Auch Stephan und Finlay (2000) konnten belegen, dass empathische Bedingungen dazu führen, dass der Bias gegenüber schwarzen im Vergleich zu weißen Zielpersonen verschwindet. Dabei kann Empathie über zwei Wege wirken: Entweder erzeugt Empathie positivere Gefühle für andere oder sie fördert die Motivation sich in unterstützender Art und Weise anderen gegenüber zu verhalten (Batson, 1991).

b) Abbau von negativem Affekt

Das „Aversive Racism Framework“ hingegen besagt, dass Vorurteile häufig auf negativen Gefühlen wie Angst und Bedrohung beruhen, die zu Unsicherheit im Umgang mit anderen und zu Vermeidung von Kontakt führen und weniger auf direkter Abneigung oder Hass gegenüber der Minderheitsgruppe (Dovidio und Gaertner, 1998). Stephan und Stephan (1985) beschreiben Intergruppenangst ebenfalls als einen wichtigen Faktor in intergruppalen Beziehungen. Weiterhin behaupten Pettigrew und Mertens (1995), dass moderne Formen gruppenbezogener

sozialer Einstellungen eher eine unterschiedliche Ausprägung positiver Emotionen als direkte Äußerungen negativer Gefühle beinhaltet. Daher sollte ihrer Ansicht nach eine Veränderung von Intergruppenangst, hin zu positiveren Gefühlen eine Veränderung in den gruppenbezogenen sozialen Einstellungen und in der Bereitschaft zu intergruppalen Kontakt bewirken (Esses & Dovidio, 2002).

c) Induktion von Schuldgefühlen

Auch gegen sich selbst gerichtete negative Gefühle können ein Mediator für wirksame Kontakt-Interventionen sein. Wenn z.B. Menschen Werte wie Gleichheit und Fairness besonders wichtig sind, erleben diese Menschen negative Gefühle, wenn sie merken, dass sie einen intergruppalen Bias zeigen und sich somit ungerecht gegenüber Mitgliedern von Minderheiten verhalten. Diese Inkonsistenzen im Selbstbild der Personen führen zu kognitiver Dissonanz, welche als unangenehm empfunden wird. Dies kann dazu führen, dass diese Menschen versuchen sich eher ohne Einstellungsbias zu äußern und zu verhalten (z.B. Devine & Monteith, 1993).

Auch Pettigrew und Tropp (2008) widmen sich in einer weiteren Metaanalyse den Wirkprozessen von Kontakt-Interventionen. Sie fassen 54 Studien mit 91 unabhängigen Stichproben, in denen zentrale Mediatoren für den Prozess hinter der Kontakthypothese geprüft werden zusammen. Die drei Mediatoren sind: 1. Wissen über die Fremdgruppe (vgl. Allport, 1954), 2. Intergruppenangst (vgl. Stephan & Stephan, 1985; Blascovich, Mendes, Hunter, Lickel & Kowai-Bell, 2001) und 3. Empathie und Perspektivenübernahme, die in dieser Metaanalyse nicht getrennt betrachtet werden (vgl. Batson, Early & Salvarani, 1997). Bei der Analyse der Zusammenhänge zeigen sich besonders die affektiven Faktoren Intergruppenangst und Empathie als bedeutsame Mediatoren für den Effekt von Kontakt auf Vorurteile. Das kognitive Wissen über die Fremdgruppe hängt zwar auch in bedeutsamer Weise mit Kontakt und Vorurteilen zusammen, allerdings sind hier die Zusammenhänge wesentlich kleiner als bei den beiden anderen Mediatoren. Die Autoren der Metaanalyse vermuten dabei eine sequentielle Abfolge in der Wirksamkeit der Mediatoren. Zunächst wird über den Intergruppenkontakt die Intergruppenangst reduziert, bevor im Anschluss die gesteigerte Empathie/Perspektivenübernahme und das Wissen über die Fremdgruppe dazu beitragen, die Vorurteile zu reduzieren. (Pettigrew & Tropp, 2008) Diese angenommene Kausalkette bedarf allerdings noch

der experimentellen Prüfung. Insgesamt können die Autoren in einem saturierten Modell (mit neun Stichproben) etwa 50 % der Kovarianz zwischen Kontakt und Vorurteilen mit den beiden Mediatoren Intergruppenangst und Empathie aufklären. Somit kommen Pettigrew und Tropp auf Grund ihrer Metaanalyse zu einem ähnlichen Schluss wie Dovidio und Kollegen (Dovidio et al., 2004) dass besonders affektive Prozesse, wie die Förderung positiver Gefühle (Empathie) oder der Abbau negativer Emotionen (u.a. Intergruppenangst) verantwortlich für die Wirksamkeit von Kontakt-Interventionen auf intergrupale Einstellungen sind. Die restliche Kovarianz wird sich über viele andere Mediatoren, wie Self-Disclosure (Turner, Hewstone & Voci, 2005), die Wahrnehmung einer größeren Variabilität in der Fremdgruppe (Islam & Hewstone, 1993) oder die gefühlte Bedrohung durch die Fremdgruppe (Tausch, Hewstone, Kenworthy, Cairns & Christ, 2007) und andere noch nicht entdeckte Prozesse erklären lassen. Gerade in dem Bereich zu Grunde liegender affektiver Prozesse scheint hier noch ein lohnendes Forschungsfeld zu liegen.

II.3.3. Stellvertretender- Kontakt-Interventionen

Die Nutzung von Kontakt-Interventionen unterliegt jedoch einer starken Einschränkung: Kontakt-Interventionen sind in vielen homogenen Umgebungen nicht durchführbar. Häufig fehlt also die nötige Fremdgruppe, um Kontakt-Interventionen zwischen Gruppen initiieren zu können. Weiterhin besteht bei diesen Interventionen immer die Gefahr, dass direkte Kontakte zwischen stark vorurteilsbehafteten Gruppen eher zu Konflikten und zur Verstärkung der Vorurteile führen.

Eine besondere Variante von Kontakt-Interventionen zur Förderung von Intergruppenbeziehungen bzw. Intergruppeneinstellungen, welche die oben genannten Probleme umgehen, sind Stellvertretender-Kontakt-Interventionen (SKI). Dieser „stellvertretende“ oder auch „erweiterte“ Kontakt findet nicht direkt zwischen Mitgliedern verschiedener Gruppen statt, sondern entsteht durch das Wissen einer Person über den Kontakt eines Eigengruppenmitglieds mit einem Fremdgruppenmitglied. Der Effekt tritt also dann auf, wenn eine Person erfährt, dass ein Mitglied der eigenen Gruppe einen engen Kontakt, oder besser noch eine Freundschaft mit einer Person aus den Reihen der Fremdgruppe pflegt. Dieses Wissen allein führt bereits zu einer positiveren Einstellung gegenüber der Fremdgruppe (Wright, Aron, McLaughlin-Volpe & Ropp, 1997).

Diese Form von Kontakt-Interventionen hat gegenüber direkten Kontakten verschiedene Vorteile, bzw. vermindert mögliche Probleme, die bei direkten Intergruppenkontakten auftreten können. Zunächst können Stellvertretender-Kontakt-Interventionen relativ leicht überall eingesetzt werden, um positivere Einstellungen gegenüber Fremdgruppen zu fördern. Es ist nicht notwendig die Zielgruppe der Intervention direkt in Kontakt mit Mitgliedern der Fremdgruppe zu bringen, womit sich diese Intervention auch zum Einsatz in z.B. ethnisch homogenen Regionen eignet. Weiterhin wird durch die Betrachtung der interagierenden Personen von außen deren Gruppenzugehörigkeit, also die spezifische Intergruppensituation, leichter salient, was für die Generalisierung von Einstellungen oder Einstellungsänderungen auf die gesamte Gruppe sehr wichtig ist (s.u. Intergruppenmodell, Hewstone & Brown, 1986). Zudem wird durch den stellvertretenden Kontakt keine Intergruppenangst hervorgerufen, die bei direkten Intergruppenkontakten auftreten kann und dann eventuell Stereotype und diskriminierendes Verhalten verstärkt (Stephan & Stephan, 1985; Wright et al., 1997).

Im Folgenden werden zunächst zwei verschiedene Formen von Stellvertretender-Kontakt-Interventionen beschrieben sowie auf mögliche Wirkprozesse eingegangen. Im Anschluss daran werden empirische Evaluationsstudien zu Stellvertretender-Kontakt-Interventionen vorgestellt.

Zwei Varianten von Stellvertretender-Kontakt-Interventionen

Es gibt zwei unterschiedliche Varianten der Vermittlung von stellvertretendem Kontakt. Zum einen die *direkte Beobachtung* von stellvertretenden Kontakten. Hier wird eine freundschaftliche Interaktion zwischen einem Eigengruppenmitglied und einem Fremdgruppenmitglied direkt beobachtet – sei es nun im Kontext einer gezielten Intervention oder durch Freunde, Bekannte oder Arbeitskollegen im normalen Alltag. Dabei ist für einen Stellvertretender-Kontakt-Effekt erforderlich, dass erstens die Gruppenzugehörigkeit der beiden Personen klar erkennbar oder bekannt ist und zweitens, dass aus der Interaktion eine Freundschaft oder zumindest eine gegenseitig freundliche Haltung deutlich wird.

Zum anderen besteht die Möglichkeit einer *medial vermittelten* Variante von Stellvertretender-Kontakt-Interventionen. Hier wird mit den Interventionsteilnehmern ein Medium, wie z.B. eine Geschichte konsumiert und aufgearbeitet, in der die Protagonisten, ein Eigengruppenmitglied und ein Fremdgruppenmitglied, gut

befreundet sind. Die Interventionsteilnehmer haben hier also keinen direkten Bezug zu dem Eigengruppenmitglied, sondern erfahren nur vermittelt über das Medium von der Intergruppenfreundschaft.

Wirkprozesse Stellvertreter- und Kontakt-Interventionen

Die Mechanismen, die für die Wirksamkeit von Stellvertretender-Kontakt-Interventionen verantwortlich sind, lassen sich nach Wright und Kollegen (1997) in drei Bereiche gliedern: Positives Beispiel eines Eigengruppenmitglieds, positives Beispiel eines Fremdgruppenmitglieds und Inclusion of others in the self.

a) Positives Beispiel eines Eigengruppenmitglieds

Eigengruppenmitglieder liefern Informationen darüber, wie eine Intergruppensituation wahrgenommen wird und in welcher Weise für die eigene Gruppe adäquat darauf reagiert wird, also wie man sich als Mitglied seiner Gruppe im Kontakt mit Mitgliedern der Fremdgruppe verhält. Dieser so genannte „referent informational influence“ (Johnston & Hewstone, 1990) wirkt besonders stark, wenn die Verhaltensnorm der Eigengruppe für eine solche Intergruppensituation noch nicht fest etabliert, ambig oder in Veränderung begriffen ist („discontinuity effect“, Schopler, Insko, Drigotas & Graetz, 1993), wie es gerade bei Kindern, die noch keinen Kontakt zu bestimmten Gruppen hatten, sein könnte. Bei bereits fest etablierten Normen kann diese durch ein positives Eigengruppenbeispiel zumindest aufgeweicht werden – besonders wenn keine negativen Konsequenzen, oder stattdessen sogar positive Sanktionen auf das positive Intergruppenverhalten folgen. Jüngere Studien konnten dazu zeigen, dass auch bei Kindern Eigengruppennormen Einfluss auf Intergruppeneinstellungen und deren Ausdruck haben (Nesdale, Durkin, Maas & Griffith, 2005; Rutland, Cameron, Milne & McGeorge, 2005).

Weiterhin kann die Beobachtung einer positiven Intergruppeninteraktion die Intergruppenangst und somit negative Einstellungen reduzieren (Islam & Hewstone, 1993; s. Kapitel II.3.2. Kontakt-Interventionen).

b) Positives Beispiel eines Fremdgruppenmitglieds

Das Beispiel eines positiven Fremdgruppenmitglieds kann den repräsentierten Prototyp der Fremdgruppe verändern. Dies geschieht besonders, wenn die Gruppenmitgliedschaft in der beobachteten Situation salient ist (Turner, Hogg,

Oakes, Reicher & Wetherell, 1987) und wenn die Fremdgruppe homogen wahrgenommen wird (Nisbett, Krantz, Jepson & Kunda, 1983; Quattrone & Jones, 1980). Weiterhin kann eine Freundschaft zwischen einem Mitglied der Fremdgruppe und einem Mitglied der Eigengruppe als Interesse und positive Einstellung der Fremdgruppe für die Eigengruppe verstanden werden und somit die Hemmung vor direktem Kontakt mit Fremdgruppenmitgliedern senken (Wright et al., 1997). Zudem werden durch die Beobachtung einer intergruppalen Interaktion auch Informationen über die Fremdgruppe und deren Kultur vermittelt. Dies macht zukünftige Missverständnisse unwahrscheinlicher und erleichtert so den Schritt zu direktem Kontakt (Triandis, 1972).

c) Inclusion of others in the self

Die Idee von Aron und Kollegen (Aron, Aron, Tudor & Nelson, 1991) lautet wie folgt: Enge Bezugspersonen werden so behandelt als wäre der Andere Teil des eigenen Selbst.

„The principle is that in a close relationship, the person acts as if some or all aspects of the partner are partially the person's own.“ (Aron et al., 1991, p. 242).

Dies beziehen die Autoren vor allem auf drei kognitive Bereiche: Ressourcen, Perspektiven und Charakteristiken. In dem Maße wie ein Partner als Teil des Selbst angesehen wird, werden Ressourcen gemeinschaftlich aufgeteilt (weil Vorteile des anderen eigene Vorteile sind), werden Akteur-Beobachter Differenzen in der Perspektive geringer (z.B. Attributionsfehler) und werden die Charakteristiken des anderen als die eigenen angesehen.

Bei stellvertretendem Kontakt werden nun Bezugspersonen aus der Eigengruppe, wie zum Beispiel Freunde, zum Teil in das eigene Selbst inkludiert (Selbst-Andere-Überlappung; Aron, Aron & Smollan, 1992). Wenn nun dieses Eigengruppenmitglied Teil seines Selbst ist und einen Freund aus der Fremdgruppe hat, der also Teil dessen Selbst ist, wird durch die Inklusion des Eigengruppenmitgliedes auch dessen Fremdgruppenfreund mit in sein Selbst aufgenommen. Somit wird also auch die Fremdgruppe zu einem gewissen Teil in das eigene Selbst inkludiert. Frei nach dem Motto „Der Freund meines Freundes ist auch mein Freund“, wird dies bei dem Stellvertretenden-Kontakt-Effekt auf die gesamte Gruppe erweitert – also: „Die Gruppe meines Freundes Freund, ist auch mein Freund“.

Empirische Studien zur Wirksamkeit von Stellvertretender-Kontakt-Interventionen

Eine britische Arbeitsgruppe um Lindsey Cameron hat sich in drei Feldstudien mit dem Thema Stellvertretender-Kontakt beschäftigt. Dabei wurden unterschiedliche Stellvertretender-Kontakt-Modelle der medial vermittelten Variante an verschiedenen Stichproben und mit unterschiedlichen Zielfremdgruppen (entweder Menschen mit Behinderung oder Flüchtlinge) auf ihre Wirksamkeit auf Vorurteile sowie Verhaltensintentionen gegenüber diesen Zielgruppen getestet. Im Folgenden werden diese Untersuchungen genauer vorgestellt.

Cameron und Rutland konnten 2006 zeigen, dass eine sechswöchige Stellvertretender-Kontakt-Intervention zu einer bedeutsamen Verbesserung der Einstellungen gegenüber Menschen mit Behinderung führt. Die Autoren haben mit 67 Kindern (27 Jungen und 40 Mädchen) im Alter zwischen fünf und zehn Jahren ($M = 8$ Jahre und 2 Monate, $SD = 9.36$ Monate) einmal wöchentlich Stellvertretender-Kontakt-Geschichten gelesen und diskutiert. Dies geschah in Kleingruppen von zwei bis drei Schülern und der Versuchsleiterin. Jede Interventionssitzung dauerte dabei etwa 20 Minuten. Die Geschichten basierten auf Unterrichtsmaterial für die entsprechende Altersstufe und waren auf das Leseniveau der teilnehmenden Kinder angepasst. Sie beinhalteten Freundschaften zwischen behinderten und nicht-behinderten Kindern, die gemeinsam Abenteuer erleben. Dabei haben die Autoren in drei Interventionsgruppen verschiedene Stellvertretender-Kontakt-Modelle gegeneinander getestet:

1. Das Dekategorisierungsmodell (Brewer & Miller, 1984) bei dem die Charaktere der Geschichten als Individuen vorgestellt werden und ihre Gruppenzugehörigkeit nicht betont wird ($n = 25$).
2. Das Intergruppenmodell (Hewstone & Brown, 1986) bei dem die Gruppenzugehörigkeit und die Typikalität der Charaktere für ihre Gruppe salient gemacht werden ($n = 22$).
3. Und eine neutrale SK-Variante ($n = 20$) bei der weder die Gruppenzugehörigkeit noch die individuellen Eigenschaften betont werden.

Je nach SK-Modell wurde der Text der Geschichten und die Gruppendiskussion im Anschluss an das Lesen der Geschichten unterschiedlich ausgestaltet und fokussiert.

Somit ergab sich für die Studie ein drei (SK-Variante: Neutral, Dekategorisierungsmodell und Intergruppenmodell) mal zwei (Messzeitpunkt: prä vs. post Intervention) mal drei (Behinderung der Zielgruppe: Lernbehinderung, körperliche Behinderung, keine Behinderung) Design. Die Teilnehmer der Studie wurden randomisiert den drei Interventionsgruppen zugeordnet. Als abhängige Effektmaße wurde eine Variante des Multi-Response Racial Attitude Measure (MRA, Doyle & Aboud, 1995) zur Erfassung von intergruppalen Einstellungen via Zuordnung von positiven und negativen Eigenschaften zu bildlich dargestellten Vertretern der jeweiligen Gruppe verwendet. Weiterhin wurde das Intended Behavior Measure (IBM, Cameron & Rutland, 2006; Abrams, Rutland & Cameron, 2003) zur Messung von Verhaltensintentionen gegenüber der jeweiligen Zielgruppe („Wie gern würdest Du mit XY spielen?“ etc. – vgl. Kapitel III.6. Messinstrumente) eingesetzt. Da sowohl die Einstellungs- als auch die Verhaltensintentionsmaße gegenüber lernbehinderten und körperlich behinderten Kindern deutlich korrelierten, wurden diese beiden Gruppen zusammengefasst, so dass es für die weiteren Auswertungen nur noch die beiden Zielgruppen behinderte Kinder (Fremdgruppe) vs. nicht-behinderte Kinder (Eigengruppe) gab.

Die Fremdgruppeneinstellung gemessen über den MRA Index (positive Einstellungen gegenüber der Fremdgruppe minus negative Einstellungen gegenüber der Fremdgruppe) veränderten sich insgesamt zum Positiven über die beiden Messzeitpunkte. Der bedeutsame Haupteffekt wird jedoch durch eine signifikante Interaktion zwischen den Faktoren Bedingung und Interviewphase erzeugt. Post-hoc zeigt sich eine signifikante Verbesserung der Fremdgruppeneinstellungen in der Interventionsgruppe mit dem Intergruppenmodell, während in den anderen beiden Gruppen keine bedeutsamen Veränderungen in den Fremdgruppeneinstellungen auftreten. Ähnliches lässt sich für die Verhaltensintentionen gegenüber der Fremdgruppe beschreiben. Der bedeutsame Haupteffekt über die zwei Messzeitpunkte wird durch die signifikante Interaktion der Faktoren Bedingung und Interviewphase spezifiziert. Allerdings erweisen sich hier zusätzlich zu den Mittelwertunterschieden der Intergruppenmodell-Interventionsgruppe auch die

Verbesserung der Dekategorisierungsmodell-Interventionsgruppe als statistisch bedeutsam. Bei den Eigengruppeneinstellungen und -verhaltensintentionen gibt es in keiner Bedingung bedeutsame Veränderungen über die Messzeitpunkte.

In einer weiteren Studie zeigten Cameron, Rutland, Brown und Douch (2006), dass sich mittels solcher Stellvertretender-Kontakt-Geschichten auch Einstellungen gegenüber ethnischen Minderheiten verändern lassen. Dazu führten die Autoren eine Studie mit 253 weißen englischen Kindern im Alter zwischen fünf und elf Jahren durch. In der Stichprobe befanden sich 116 Jungen und 137 Mädchen. Auch in dieser Studie testeten die Autoren verschiedene Modelle der Stellvertretender-Kontakt-Intervention zur Förderung von positiven Intergruppeneinstellungen gegenüber Flüchtlingen.

1. Das Dekategorisierungsmodell (Brewer & Miller, 1984) bei dem die Charaktere der Geschichten individuell vorgestellt werden und ihre Gruppenzugehörigkeit nicht betont wird (s.o.) ($n = 70$).
2. Das Common Ingroup Identity Modell (Gaertner, Mann, Murrell & Dovidio, 1989) bei dem betont wird, dass die Charaktere der Geschichten einer gemeinsamen übergeordneten Kategorie, hier der gleichen Schule, angehören ($n = 68$).
3. Das Dual Identity Modell (Gaertner & Dovidio, 2000; Gonzalez & Brown, 2003) bei dem sowohl eine gemeinsame übergeordnete Gruppe, die gleiche Schule, der Charaktere als auch deren Zugehörigkeit zu ihren unterschiedlichen ethnischen Subkategorien betont wird ($n = 69$).
4. Kontrollgruppe ($n = 54$)

Auch hier wurden die Teilnehmer an der Untersuchung randomisiert den einzelnen Bedingungen zugewiesen. Die Intervention verlief ähnlich wie in der oben geschilderten Studie mit der Ausnahme, dass die verwendeten Abenteuergeschichten dieses Mal von englischen weißen Kindern handelten, die mit Flüchtlingskindern aus anderen Ländern befreundet sind.

In dieser Studie untersuchten die Autoren zusätzlich zur Wirksamkeit der verschiedenen SK-Modelle noch den Einfluss des Alters, indem sie die Kinder in jüngere, fünf bis acht Jahre ($M = 7$ Jahre und 2 Monate, $SD = 5.96$ Monate), und

ältere, neun bis elf Jahre ($M = 10$ Jahre und 6 Monate, $SD = 7.12$ Monate), einteilten. Ebenso erfassten die Autoren den Einfluss der Identifikation mit der Eigengruppe (auch hier wurde ein dichotomisierter Faktor hoch vs. tief Identifizierer errechnet), sowie den möglichen Mediator Inclusion of others in the self. Abhängige Maße waren abermals die Einstellungen der Kinder sowie die Verhaltensintentionen gegenüber Mitgliedern der Zielgruppe. Diese wurden über eine Abwandlung des Multi-Response Racial Attitude Measure und das Intended Behavior Measure (s.o.) ausschließlich nach der Interventionsphase erfasst. Somit ergab sich für diese Studie ein vier (Bedingung: Dekategorisierungsmodell, Common Ingroup Modell, Dual Identity Modell, Kontrollgruppe) mal zwei (Altersgruppe: 5-8 Jahre vs. 9-11 Jahre) mal zwei (Identifikation (englisch): hoch vs. tief) Design. Die Ergebnisse der statistischen Analysen zeigen, dass die Fremdgruppeneinstellungen der Kinder in allen Stellvertretender-Kontakt-Bedingungen signifikant positiver ausfallen als in der Kontrollbedingung. Dabei hebt sich das Dual Identity Modell nochmals von den anderen beiden SK-Bedingungen positiv ab und erweist sich hier als wirksamer als das Common Ingroup Modell oder das Dekategorisierungsmodell. Allerdings konnten die Autoren keine Unterschiede zwischen den Bedingungen im Bereich des intendierten Verhaltens gegenüber Fremdgruppenmitgliedern finden. Der Faktor Alter wirkt sich auf allen abhängigen Maßen aus. Die jüngeren Kinder zeigen sowohl bei den Einstellungen als auch bei den Verhaltensintentionen signifikant positivere Werte der Fremdgruppe gegenüber als die älteren Kinder. Dies gilt im Übrigen auch für die Maße zur Eigengruppe. Der Faktor Identifikation mit der Eigengruppe trägt hauptsächlich zur Aufklärung von Einstellungen und Verhaltensintentionen gegenüber der Eigengruppe bei. Die hoch-Identifizierer zeigen sich positiver in Einstellungen und Verhaltensintentionen gegenüber der Eigengruppe als die tief-Identifizierer. Weiterhin konnten Cameron und Kollegen in dieser Studie Inclusion of others in the self als Mediator für den Effekt der Intervention auf die Intergruppeneinstellung via Regressionsanalysen bestätigen.

Hier zeigte sich also das Dual Identity Modell (Gaertner & Dovidio, 2000) als besonders erfolgreich in der Beeinflussung von Einstellungen. Allerdings reichen die Effekte der Intervention nicht bis in den Bereich der Verhaltensintentionen.

In zwei weiteren Studien der Arbeitsgruppe um Lindsey Cameron (Cameron, Rutland & Brown, 2007) konnten die positiven Ergebnisse der ersten beiden Studien

repliziert werden. Hier ging es abermals um den Effekt von Stellvertretender-Kontakt-Interventionen auf die Einstellungen und Verhaltensintentionen gegenüber Kindern mit Behinderung. Allerdings wurde hier die Wirksamkeit der SK-Intervention mit der Wirksamkeit einer anderen Interventionsart verglichen: Dem Training der multiplen Klassifikationsfähigkeit (vgl. Bigler & Liben, 1992). In der ersten Studie wurden $N = 71$ Kinder (35 Jungen und 36 Mädchen) im Alter zwischen sieben und neun Jahren ($M = 7$ Jahre und 9 Monate, $SD = 8.66$ Monate) auf drei Gruppen aufgeteilt: Die Kontrollgruppe ($n = 28$), eine Stellvertretender-Kontakt-Intervention ($n = 21$) und ein Training der multiplen Klassifikationsfähigkeit ($n = 22$). Die angewendete Stellvertretender-Kontakt-Intervention verlief dabei wie die Intervention in der Studie von Cameron und Kollegen (s.o., Cameron et al., 2006a), wobei das erfolgreiche Intergruppenmodell (Hewstone & Brown, 1986) auch hier angewandt wurde.

Das Training der multiplen Klassifikationsfähigkeit basiert auf dem Training das Bigler und Liben entwickelt und getestet haben (Bigler & Liben, 1992) haben. Es entsprach in Dauer und Intensität den Sitzungen der SK-Intervention (über 6 Wochen jeweils eine Sitzung à 20 Minuten). Bei diesem Training übten die Kinder, soziale Stimuli mit unterschiedlichen Merkmalsdimensionen (körperlich behindert/lernbehindert/nicht behindert, glücklich/unglücklich, hilfsbereit/nicht hilfsbereit, mag Computer/mag Bücher) in ein Vierfelderschema ein zu sortieren. Die Teilnehmer erhielten dazu zwölf Photographien von behinderten und nicht behinderten Kindern auf denen zusätzlich Symbole abgedruckt waren, welche Informationen über die übrigen Merkmalsausprägungen der abgebildeten Kinder darstellten. Die Kinder sollten die Bilder zunächst nach einer Dimension sortieren (Behinderung) dann nach einer zusätzlichen Dimension (glücklich/unglücklich) und in jeder weiteren Sitzung nach einer anderen weiteren der oben genannten Dimensionen. Außerdem sollten die Kinder ihre Sortierung begründen. Bei falscher Sortierung wurden die Kinder vom Versuchsleiter korrigiert.

Als abhängige Maße wurden hier eine Variante des Multi-Response Racial Attitude Measure sowie das Intended Behavior Measure in Bezug auf die Gruppen körperlich behinderte, lernbehinderte und normale Kinder verwendet (s.o.). Auch in dieser Studie wurden die teilnehmenden Kinder nur einmal befragt, und zwar eine Woche nach dem Ende der Interventionen.

In den Ergebnissen der Untersuchung zeigte sich weder ein Einfluss des Geschlechts der Probanden noch ein Unterschied in der Beurteilung von körperlich

behinderten oder lernbehinderten Kindern. Aus diesem Grund wurde der Faktor Geschlecht aus den weiteren Analysen ausgeschlossen und die korrelierenden Einstellungen sowie die Verhaltensintention gegenüber den beiden Fremdgruppen zu einer Fremdgruppeneinstellung und einer Fremdgruppenverhaltensintention zusammengefasst.

Die Ergebnisse der Untersuchung zeigten, dass die Kinder in der SK-Bedingung eine signifikant positivere Einstellung gegenüber der Fremdgruppe als die Kinder in den beiden anderen Bedingungen hatten. Die Verhaltensintention gegenüber der Fremdgruppe der SK-Teilnehmer war ebenfalls positiver im Vergleich zu den Teilnehmern des multiplen Klassifikationstrainings, allerdings nicht positiver als die der Kontrollgruppe. In den Maßen der Eigengruppeneinstellung und Eigengruppenverhaltensintention fanden sich keine bedeutsamen Mittelwertsunterschiede zwischen den Experimentalbedingungen.

In einer weiteren Folgestudie testeten die gleichen Autoren die Wirkung dreier verschiedener Interventionsarten auf die Einstellungen und Verhaltensintentionen gegenüber Flüchtlingen. Die einzelnen Interventionsformen waren dabei Stellvertretender-Kontakt ($n = 51$), wie er in der ersten Studie eingesetzt wurde, nur hier mit Flüchtlingen anstatt mit Menschen mit Behinderung als Fremdgruppe. Ein überarbeitetes Training der multiplen Klassifikationsfähigkeit ($n = 55$) mit Bildern von weißen und nicht-kauasischen Erwachsenen die mit zusätzlichen Symbolen verschiedener Merkmalsdimensionen wie in der ersten Studie versehen waren. Dazu wurde noch die Dimension des statushohen vs. statusniedrigen Berufs eingeführt. Ansonsten verlief das Training der multiplen Klassifikation wie in Studie eins. Weiterhin gab es eine Gruppe von Teilnehmern, die die Kombination aus beiden Interventionen in vollem Umfang erhielten ($n = 45$). Zusätzlich gab es auch hier eine Kontrollgruppe ($n = 47$). Die Stichprobe umfasste insgesamt 198 Kinder (99 Jungen und 99 Mädchen) im Alter von sechs bis elf Jahren. Neben dem Effekt der Bedingung wurde auch die Wirkung des Alters der Kinder mit jüngeren Kindern im Alter von sechs bis acht Jahren ($M = 7$ Jahre und 10 Monate, $SD = 8.76$ Monate) im Vergleich zu älteren Kindern im Alter von neun bis elf Jahren ($M = 10$ Jahre und 4 Monate, $SD = 7.74$ Monate) untersucht. Ebenso wurde die Identifikation der Kinder mit ihrer Nationalität erfasst und als dichotomisierter Faktor (Identifikation hoch vs. Identifikation niedrig – basierend auf dem Mediansplit) auf ihren Einfluss hin überprüft.

Somit ergab sich für diese Studie ein vier (Bedingung: Stellvertretender-Kontakt, Training multipler Klassifikation, Kombination aus beidem, Kontrollgruppe) mal zwei (Englische Identifikation hoch vs. niedrig) mal zwei (Altersgruppe: jüngere Kinder vs. ältere Kinder) Design. Als abhängige Effektmaße kamen eine Variante des Multi-Response Racial Attitude Measure sowie das Intended Behavior Measure (vgl. Cameron et al., 2006a) zum Einsatz. Weiterhin wurde die multiple Klassifikationsfähigkeit mittels eines Klassifikationstest nach Bigler (1995) erfasst und schließlich auch hier die Identifikation mit englischen Kindern mittels einer kurzen Skala von Verkuyten (2001) gemessen.

Die Ergebnisse der Folgestudie replizieren die Ergebnisse der ersten Studie. Die Experimentalbedingung hat einen bedeutsamen Einfluss auf die gemessene Fremdruppeneinstellung. Hier zeigt sich post-hoc, dass die Bedingungen mit Stellvertretender-Kontakt-Interventionen eine positivere Fremdruppeneinstellung aufweisen als die mit multiplen Klassifikationstraining oder in der Kontrollgruppe. Bei den Verhaltensintentionen gegenüber der Fremdgruppe zeigt sich ein signifikanter Alterseffekt. Jüngere Kinder zeigen sich gegenüber Fremdgruppenmitgliedern generell positiver in ihren Verhaltensintentionen als ältere Kinder. Weiterhin findet sich noch eine Interaktion zwischen den Faktoren Bedingung und Identifikation: die hoch identifizierten Kinder haben in den Stellvertretender-Kontakt-Bedingungen eine positivere Verhaltensintention gegenüber der Fremdgruppe als die niedrig identifizierten Kinder.

Ansonsten gab es im Bereich der Eigengruppeneinstellungen wie in den anderen Studien einen Alterseffekt sowie einen Effekt der Identifikation. Jüngere Kinder beurteilten ihre Gruppe positiver als ältere und hoch identifizierte Kinder hatten eine positivere Einstellung zur Eigengruppe als niedrig identifizierte. Ein ähnlicher Alterseffekt fand sich bei den Verhaltensintentionen gegenüber der Eigengruppe. Und nicht zuletzt fand sich eine Interaktion zwischen dem Alter der Kinder und der Experimentalbedingung: Die älteren Kinder hatten in der kombinierten Interventionsbedingung eine positivere Eigengruppenverhaltensintention als die Kontrollgruppe. Die Fähigkeit zur multiplen Klassifikationsfähigkeit war in den Gruppen die ein entsprechendes Training erhielten signifikant besser als in der reinen SK-Bedingung und in der Kontrollgruppe. Allerdings erwies sich die Klassifikationsfähigkeit in einer Regressionsanalyse nicht als bedeutsamer Prädiktor

der abhängigen Maße und hatte also keinen Einfluss auf die hier gemessenen Einstellungen und Verhaltensintentionen der Untersuchungsteilnehmer.

Zusammenfassung und Fazit zu Stellvertretender-Kontakt-Interventionen

Insgesamt erweist sich Stellvertretender-Kontakt als eine erfolgreiche Methode um positive Fremdgruppeneinstellungen bei Kindern zu fördern und dies über einen breiten Altersbereich von sechs bis elf Jahren. Dabei scheint das Intergruppenmodell von Hewstone und Brown (1986, 2005) besonders effektiv zu sein. Der Erfolg des Dual Identity Ansatzes in der Studie von Cameron und Kollegen (2006b) scheint eher auf die hier betonte Zugehörigkeit der SK-Charaktere zu ihren Subgruppen zurückzuführen zu sein statt auf die gemeinsame übergeordnete Kategorie. Dafür sprechen die Ergebnisse aus den anderen drei Untersuchungen (Cameron & Rutland, 2006b; Cameron et al., 2007), in denen das Intergruppenmodell am Besten funktioniert hat, während das Common Ingroup Identity Modell nicht so erfolgreich war (Cameron, Rutland, Brown & Douch, 2006). Wobei angemerkt werden muss, dass die beiden Modelle (Intergruppenmodell und Dual Identity Modell) nie in einer Studie gegeneinander getestet wurden. Die Salienz der Gruppenzugehörigkeit sowie die Typikalität der SK-Charaktere scheinen aber für die SK-Wirkung auf die Fremdgruppeneinstellungen besonders wichtig zu sein.

Im Vergleich mit einer anderen Interventionsform, dem Training multipler Klassifikationsfähigkeit, zeigte sich in den Studien von Cameron und Kollegen allein die Stellvertretender-Kontakt-Intervention in der Intergruppenmodellvariante als wirksam in der Veränderung von Einstellungen gegenüber Menschen mit Behinderung und Flüchtlingen bei Kindern. Wobei hier festgehalten werden muss, dass Bigler und Liben einen Einfluss der multiplen Klassifikationsfähigkeit auf Fremdgruppenstereotype und nicht auf stärker affektiv getönte Fremdgruppeneinstellungen aufzeigen konnten (Bigler & Liben, 1992; Cameron et al., 2007). Somit kann dies als Hinweis für eine getrennte Betrachtung dieser beiden Konstrukte eingeordnet werden (vgl. Augoustinos & Rosewarne, 2001; Devine, 1989).

Allerdings weisen die Studien der Arbeitsgruppe um Lindsey Cameron auch einige methodische Mängel auf. So bleibt das positive Ergebnis der ersten Studie (Cameron & Rutland, 2006) fragwürdig, weil hier keine vergleichende Kontrollgruppe erfasst wurde. Die Teilnehmer wurden zwar randomisiert zu den

Experimentalbedingungen zugewiesen, das verhindert jedoch nicht zwangsläufig, dass es systematische konfundierende Unterschiede zwischen den Gruppen geben kann, wie zum Beispiel in den Prätestwerten der Fremdgruppeneinstellungen (insbesondere bei kleineren Stichproben wie bei Cameron et al., 2006a). In den anderen Untersuchungen wurde zwar mit einer Kontrollgruppe gearbeitet, dafür jedoch auf einen Prätest verzichtet. Hier sei darauf hingewiesen, dass eine randomisierte Zuweisung eine Konfundierung der abhängigen Maße zwar unwahrscheinlicher macht, aber nicht ausschließt. Weiterhin ist bedauerlich, dass die Autoren auf die Angabe von Effektstärken verzichten, die zumindest querschnittlich im Vergleich zu Kontrollgruppe bzw. in der Arbeit von Cameron und Rutland (2006) auch längsschnittlich ohne Bezug zur vergleichenden Kontrollgruppe hätten berechnet werden können, um die praktische Bedeutsamkeit der Stellvertretender-Kontakt-Interventionen einschätzen zu können. Nicht zuletzt fehlen in der Wirksamkeitsprüfung der Intervention die Evaluation langfristiger Interventionseffekte sowie eine Implementationskontrolle. Die Messung der abhängigen Maße fand immer eine Woche nach Ende der Intervention statt. Follow-ups wurden nicht durchgeführt. Somit bleibt völlig offen, ob die Stellvertretender-Kontakt-Intervention auch langfristig wirksam ist oder nicht.

II.3.4. Wissensbasierte Interventionen

Wissensbasierte Interventionen sind ein weit verbreitetes und traditionsreiches Verfahren zur Reduktion von Vorurteilen und Diskriminierung. Die Grundlage dieser Interventionsprogramme ist die Vermittlung von Wissen über verschiedene Gruppen. Dabei wird häufig auf die Geschichte, die Kultur, die Lebensumstände sowie auf Normen und Werte der jeweiligen Gruppe eingegangen. Dieses Wissen über eine Fremdgruppe soll dann zu positiveren Einstellungen und Verhaltensweisen gegenüber dieser Gruppe führen.

Programme die wissensbasiert ansetzen, sind jedoch selten reine wissensbasierte Interventionen. Sie beinhalten häufig zusätzlich andere Elemente und werden unter dem breiten Begriff der Multicultural Education zusammengefasst. Gemeinsam haben diese Programme, dass mit ihrer Hilfe versucht wird, Schülern direkt etwas über verschiedene ethnische, kulturelle und religiöse Gruppen in ihrer Gesellschaft beizubringen. Weitere Programminhalte sind dabei häufig soziale Schichten, aber auch Informationen über die Geschichte der einzelnen Gruppen, sowie über deren kulturelle Werte und Praktiken; und Informationen über deren Alltag. Es finden sich zusätzlich häufig Elemente in den Programmen, die sich mit der Kultur der Eigengruppe und den individuellen Stärken und Schwächen der Teilnehmer auseinandersetzen. Hier soll oft der Selbstwert der Teilnehmer gefördert werden – mit der Idee, dass Personen mit positivem Selbstwert „es nicht nötig haben“ andere abzuwerten. Außerdem sind Themen wie Vorurteile, Stereotype und Diskriminierung ebenfalls oft Bestandteile von Multicultural Education Programmen. Dabei herrscht häufig ein direkter didaktischer Ansatz vor, der durch interaktive Elemente wie Rollenspiele, Gruppendiskussion oder Simulationsspiele erweitert wird, um die Kinder als aktive Teilnehmer besser in das Programm einzubinden und somit die Wirksamkeit des Programms zu erhöhen (s. Kapitel II.3.1. Klassifikation von Interventionsansätzen).

Die Ziele von Multicultural Education Programmen basieren im Allgemeinen auf einem Set demokratischer Werte wie Gleichheit von Möglichkeiten, Gerechtigkeit für alle, Menschenwürde, Respekt vor Gruppenunterschieden und Pluralismus (Stephan & Stephan, 2001). Diese Werte sowie angemessene soziale und persönliche Standards und Normen zu Akzeptanz und Inklusivität sollen in zahlreichen Programmen an die Teilnehmer vermittelt werden.

Ein Beispiel einer derartigen Intervention ist das Programm „Eine Welt der Vielfalt“ (1998) des „A World of Difference-Instituts“ der Anti-Defamation League aus den USA. In diesem Programm gibt es viele verschiedene Übungen und Spiele aus fünf Bereichen: Wertschätzung von sich selbst und anderen als Mitglied verschiedener Gruppen, Ähnlichkeiten und Unterschiede zwischen Gruppen, kulturelle Vielfalt, das Wesen und die Auswirkungen von Vorurteilen, Strategien zur Bekämpfung von Vorurteilen. Die verschiedenen Übungen sind dabei für unterschiedliche Altersbereiche zugeschnitten und eignen sich für Kinder der Primarstufe sowie der Sekundarstufe I.

Wirkprozesse wissensbasierter Interventionen

Es werden zahlreiche und zum Teil sehr unterschiedliche Prozesse der Wirksamkeit von wissensbasierten Interventionen vermutet. Dovidio und Kollegen (2004) befassen sich ausführlich mit Wirkprozessen von Interventionen zur Reduktion negativer gruppenbezogener sozialer Einstellungen. Im Gegensatz zu Kontakt-Interventionen (s.o.), nehmen Dovidio und Kollegen für wissensbasierte Interventionen (Aufklärung über Kulturen, Vorurteile und deren Wirkung, Verantwortung für sozialen Wandel etc) an, dass bei solchen aufklärenden Interventionen eher kognitive Wirkprozesse zu Grunde liegen. Allerdings greifen die unterschiedlichen Prozesse oft ineinander, so dass eine klare Trennung nicht einfach ist. Die Prozesse, die von Dovidio und Kollegen in diesem Zusammenhang beschrieben werden, werden im Folgenden näher erläutert.

a) Soziale Kategorisierung

Soziale Kategorisierung und besonders die Einteilung von Personen in Eigen- oder Fremdgruppe führt meist zu einer Verzerrung in Bezug auf Bewertungen, Attributionen, Ressourcenverteilung, Hilfeverhalten und soziale Unterstützung gegenüber diesen Personen (Mullen, Brown & Smith, 1992; Messick & Mackie, 1989). Kategorisierungsprozesse können aber genutzt werden, um solche Verzerrungen zu verringern. Entweder über Dekategorisierungsprozesse, ausgelöst durch die Vermittlung von individuellen Informationen wodurch Menschen eher als Individuum betrachtet werden und weniger als Mitglied einer Gruppe; oder über Rekategorisierungsprozesse, wobei Personen als Mitglied einer gemeinsamen übergeordneten Kategorie betrachtet werden und somit als Mitglied der eigenen

meist positiv bewerteten Gruppe. Die Schaffung einer solchen übergeordneten Kategorie und die Betonung übergeordneter sozialer Identitäten wie gemeinsamer Wohnort oder Nationalität kann die Wahrnehmung von Gruppenunterschieden sowie Vorurteile senken (Dovidio, Kawakami & Gaertner, 2000). Die Betonung solcher so genannter „cross-cutting identities“, also sich in verschiedenen Bereichen überschneidenden Identitäten (z.B. Fan des gleichen Fußballclubs, Bewohner des gleichen Stadtviertels etc.), kann dazu beitragen, dass die wahrnehmungsverzerrende Wirkung von unterschiedlichen Gruppenzugehörigkeiten relativiert wird (Stephan & Stephan, 2001).

b) Soziales Wissen und Wahrnehmung von Gruppen

Der Erwerb von Wissen über Menschen anderer Gruppen kann die soziale Wahrnehmung dieser Gruppen und die Wahrnehmungsverzerrungen ihnen gegenüber auf verschiedene Arten verringern:

Mehr Wissen über andere führt dazu, dass diese als individualisierter wahrgenommen werden, was die Notwendigkeit auf stereotypes Wissen zurückzugreifen und somit negative Wahrnehmungsverzerrungen reduziert

Mehr Wissen über Andere reduziert die Unsicherheit im Umgang mit diesen Anderen, und kann zur Minderung von Ignoranz und Angst gegenüber dieser Gruppe beitragen. Bedenken vor einer Interaktion mit Fremdgruppenmitgliedern können so zerstreut werden. Dies kann zu größerem Verständnis und Akzeptanz führen und verringert somit ebenfalls die Vermeidung von intergruppalen Kontakt sowie unangenehme Gefühle, die bei intergruppalen Kontakt entstehen können (Crosby, Bromley & Saxe, 1980; Gaertner & Dovidio, 1986). Positiver intergruppaler Kontakt wiederum verringert Vorurteile und Diskriminierung (s. Kapitel II.3.2. Kontakt-Interventionen).

Mehr Wissen über andere (gerade über historische Hintergründe) kann den intergruppalen Bias reduzieren, indem die Behandlung der Fremdgruppe als ungerecht empfunden wird (Stephan & Finlay, 1999) und indem Empathie gegenüber der Fremdgruppe ermöglicht und gefördert wird. Gerade Material das aus der Sicht eines Fremdgruppenmitgliedes dargeboten wird, kann dazu dienen die Perspektive von Fremdgruppenmitgliedern zu übernehmen und somit zu mehr Empathie mit diesen Fremdgruppenmitgliedern führen. Empathie wiederum kann Vorurteile reduzieren (Batson et al., 1997; Stephan & Finlay, 2000).

Mehr Wissen kann negative gruppenbezogene soziale Einstellungen aber auch reduzieren, indem die Beziehung zwischen den Gruppen und ihren Mitgliedern anders betrachtet wird. Statt einer kompetitiven Beziehung sollte eine kooperative und interdependente Beziehung zwischen den Gruppen wahrgenommen werden – dies reduziert ebenfalls die intergruppalen Wahrnehmungsverzerrungen (Sherif et al., 1961; Campbell, 1965; Bobo, 1988; Esses, Dovidio, Jackson & Armstrong, 2001; Sidanius & Pratto, 1999).

c) Verhaltensnormen und Werte

Wissen kann ebenfalls negative gruppenbezogene soziale Einstellungen reduzieren, indem auf Grund von neuem Wissen die persönlichen Verhaltensnormen neu bewertet und verändert werden. Dieses Wissen kann die Wahrnehmung, was angemessen und gerecht ist, verändern und somit die Konstruktion oder Rekonstruktion sozialer Normen beeinflussen. Der Einfluss von Normen auf Einstellungen und Vorurteile wurde gut untersucht (Duckitt, 1992). Dabei kann zwischen persönlichen und sozialen Normen unterschieden werden (Allport, 1954; Devine, 1989). Macrae, Bodenhausen und Milne (1998) konnten dazu zeigen, dass Personen, die sich auf sich selbst konzentrieren, also mehr an ihren persönlichen Normen und Standards orientiert sind, andere weniger stereotyp beschreiben, zumindest wenn ihren persönlichen Standards eine stereotype Denkweise zuwider läuft.

Außerdem konnten Wittenbrink und Henly (1996) zeigen, dass wahrgenommene Normen und Standards über den Umgang der Eigengruppe mit Fremdgruppen das Verhalten von Personen beeinflussen können. So konnten sie zeigen, dass weiße Studenten weniger Vorurteile in einer Geschworenenaufgabe in einem Gerichts-Szenario gegenüber einem schwarzen Angeklagten zeigten, wenn sie zuvor die Information erhielten, dass Weiße insgesamt eher wenig vorurteilsbelastet gegenüber Schwarzen sind (s.o. referent informational influence (Johnston & Hewstone, 1990)). Wenn Schüler also positive humanitäre Werte verinnerlicht haben und die Norm gelernt haben, dass Vorurteile falsch sind, werden sie weniger auf Vorurteile und Stereotype zurückgreifen und kaum diskriminierendes Verhalten mehr zeigen (Devine, 1995).

Insgesamt kann festgehalten werden, dass kognitive Prozesse, welche die Kategorisierung, die Wahrnehmung von Fremdgruppen sowie angemessene Standards über den Umgang mit der Fremdgruppe und den Ausdruck von sozialen Einstellungen beeinflussen, gruppenbezogene soziale Einstellungen senken können.

Kognitive und affektive Mediatoren in der Reduzierung negativer gruppenbezogener sozialer Einstellungen treten jedoch oft miteinander verknüpft auf. Zum Beispiel wird durch das Fokussieren auf persönliche Standards von Freiheit und Gleichheit die individuelle Betrachtung von Personen bekräftigt (kognitiver Prozess) und gleichzeitig werden aber eventuell auch Schuldgefühle über eigene Vorurteile erzeugt (affektiver Prozess).

Abschließend sollte bei Interventionen bedacht werden, dass affektiv fokussierte Prozesse sich eher dazu eignen affektive Einstellungsformen (z.B. negativen Affekt) zu beeinflussen, während kognitiv orientierte Interventionen eher kognitive Einstellungsformen (Stereotype, Eigenschaftszuschreibungen) verändern können (Dovidio et al., 2002; Esses & Dovidio, 2002).

Empirische Befunde zur Wirksamkeit wissenschaftlicher Interventionen

Im Folgenden werden einige für die vorliegende Arbeit zentrale Studien mit wissenschaftlichen Interventionen dargestellt. Zunächst wird kurz auf einige ältere Studien eingegangen, um dann zwei aktuellere Studien ausführlicher vorzustellen und schließlich noch zwei Überblicksarbeiten zu diesem Thema zu präsentieren.

Ältere Evaluationsstudien zu wissenschaftlichen Interventionen

Bereits in den vierziger und fünfziger Jahren fanden die ersten Untersuchungen zur Wirkung von Unterrichtsmaterialien auf die gruppenbezogenen sozialen Einstellungen von Kindern statt. So untersuchten schon Jackson (1944) und Agnes (1947) den Effekt von Unterrichtsmaterial auf die ethnischen Einstellungen von Schülern. Dabei zeigte sich, dass Schüler, welche Material über Afroamerikaner gelesen hatten, eine positivere Einstellung gegenüber Schwarzen hatten, als die Schüler, die nichts über Afroamerikaner gelesen hatten. Allerdings ließen die Effekte schnell wieder nach und die Studien beinhalteten schwerwiegende methodische Mängel, was die Generalisierbarkeit der Effekte deutlich einschränkt (Banks, 2001).

Trager und Yarrow (1952) untersuchten die Wirkung eines demokratischen Curriculums versus ein normales amerikanisches Curriculums auf die ethnischen Einstellungen von Erst- und Zweitklässlern. Die curriculare Intervention bestand aus Texten, Aktivitäten und Lehrern, die entweder einen demokratischen oder einen nicht-demokratischen Lehrstil praktizierten. Nach der Intervention zeigten die Schüler, die ein demokratisches Curriculum besuchten, positivere ethnische Einstellungen und Verhaltensweisen als die Schüler, die das herkömmliche Curriculum besuchten. Interessanterweise zeigten nach der Intervention auch die Lehrer in der Experimentalbedingung positivere Einstellungen gegenüber Fremdgruppen.

Fisher (1965) konnte zeigen, dass das Lesen und besonders das Lesen und Diskutieren von sechs Geschichten über amerikanische Indianer eine positivere Einstellung zu dieser Gruppe bei Fünftklässlern im Vergleich zur Kontrollgruppe bewirkt. Die Einstellungsveränderung war in der Experimentalgruppe mit anschließender Diskussion über die Geschichten signifikant größer als in der Gruppe, in der die Geschichten nur gelesen wurden. Auch Tauran (1967) konnte zeigen, dass das Lesen von positiver aber auch negativer Literatur die Einstellung von Drittklässlern gegenüber Eskimos in die entsprechende Richtung beeinflusst. Natürlich ist eine derartige Studie ethisch bedenklich, da einem Teil der Kinder negative Informationen und Einstellungen gegenüber Eskimos nahe gelegt wurden. Allerdings weist diese Studie darauf hin, dass Interventionen und Interventionsmaterial sorgfältig geplant, gesichtet und durchgeführt werden müssen, um derartige negative Effekte zu vermeiden.

Yawkey (1973) testete die Wirkung von Informationen über die Geschichte, die Kultur und die soziale Situation von Afroamerikanern auf die Einstellungen von weißen Schülern gegenüber dieser Fremdgruppe. Dazu wurden im Unterricht sechs Bücher über die Kultur und Geschichte von Afroamerikanern gelesen und unter der Leitung der Lehrer diskutiert. Die Intervention hatte einen positiven Effekt auf die Einstellungen der Schüler gegenüber Afroamerikanern.

Litcher und Johnson (1969) testeten die Wirksamkeit von Unterrichtsmaterialien und Büchern mit Bildern, in denen schwarze Charaktere aus der Mittelklasse dargestellt wurden, wie sie mit weißen Personen aus der Mittelklasse interagieren. Ihre Stichprobe bestand aus weißen Zweitklässlern ohne viel Kontakt zu schwarzen Personen. Die Kontrollgruppe erhielt die normalen Unterrichtsmaterialien

mit Darstellungen von hauptsächlich weißen Personen. Das Programm verlief über vier Monate. Am Ende der Interventionszeit zeigten die Kinder mit den „schwarzen“ Unterrichtsmaterialien eine signifikant positivere Einstellung gegenüber Schwarzen auf allen vier eingesetzten Instrumenten: Dem Clark Doll Test (Gregor & McPherson, 1966), dem Show Me Test (Horowitz & Horowitz, 1938), dem Categories Test (Horowitz & Horowitz, 1938) und dem Direct Comparison Test (Blake & Dennis, 1943). Bei dieser Interventionsform eröffnet sich allerdings die Frage, ob derartige Materialien wirklich als wissensbasierte Intervention zu sehen sind oder ob hier nicht auch Prozesse des stellvertretenden Kontakts (s. Kapitel II.3.3. Stellvertretender-Kontakt-Interventionen) eine Rolle spielen.

Unabhängig davon konnten Litcher, Johnson und Ryan diese positiven Ergebnisse in einer Folgestudie (1973) nicht replizieren. Hier untersuchten sie allerdings eine Stichprobe von Zweitklässlern, welche in einer weniger homogenen Umgebung aufwuchsen und schon viel Kontakt zu schwarzen Personen hatten. Somit war die Einstellung der Kinder zu Schwarzen vermutlich schon deutlich elaborierter und gefestigter als bei den Teilnehmern der ersten Studie, die aus relativ homogen weißen Gegenden und Schulen stammten. Außerdem verkürzten die Autoren die Interventionsdauer auf einen Monat und veränderten das Interventionsmaterial, das nun weniger Bilder beinhaltete. Leider lässt sich durch diese verschiedenen Variationen in der Replikationsstudie nicht genau feststellen, welche Faktoren für die Wirksamkeit bzw. Unwirksamkeit der Intervention verantwortlich sind. Es könnte sowohl an der Dauer als auch an den Kontakterfahrungen der Teilnehmer oder den veränderten Materialien gelegen haben.

Die Wirkung des Lesens von Geschichten über Amerikaner mexikanischer Herkunft untersuchte Howell (1973). Zusätzlich variierte er die Intensität und Dauer in seinen beiden Experimentalbedingungen: Eine Gruppe las zehn Geschichten über zehn Wochen verteilt, die andere Gruppe las fünf Geschichten über neun Wochen verteilt. Die Kontrollgruppe las in der Zeit keine Geschichten über die Zielgruppe. Die Gruppe mit zehn Geschichten zeigte nach der Intervention eine signifikant positivere Einstellung auf drei von vier Einstellungsmaßen, während die zweite Gruppe mit den fünf Geschichten lediglich auf einem Maß eine bedeutsame Verbesserung ihrer Einstellungen gegenüber Mexikoamerikanern erzielen konnte. Diese Studie macht

deutlich, dass Intensität einer wissensbasierten Intervention entscheidend für ihre Wirksamkeit sein kann³.

Lessing und Clarke (1976) hingegen fanden keine positiven Effekte einer achtwöchigen Intervention zur Förderung von Intergruppenbeziehungen bei Junior-Highschool-Schülern. Die Intervention beinhaltete das Lesen interkultureller Literatur, die Einladung von Gastrednern und die Vorbereitung eigener Referate zu dem Thema. Problematisch an der Studie war, dass die Teilnehmer sich die Bücher, welche sie lesen sollten, selbst aussuchen konnten, und dass diese Bücher von fünf verschiedene Ethnien handelten. Somit erhielten die Schüler in der Experimentalgruppe keine standardisierte Intervention. Außerdem fand die Studie in einer fast homogen weißen Schule und Gemeinde statt, in der vermutlich ungünstige Einstellungen und Beziehungen zwischen den Ethnien vorherrschten, so dass die Bücher, welche die Schüler gelesen hatten, die negativen Gefühle der Schüler den Minderheiten gegenüber eventuell noch verstärkt haben (Lessing & Clarke, 1976). Dies zeigt, dass nicht jede wissensbasierte Intervention eine positive Wirkung entfalten muss.

Westphal (1977) untersuchte die Wirkung eines interkulturellen Curriculums auf die Einstellungen von weißen Erstklässlern (das Curriculum wurde bereits in der Vorschule durchgeführt). Das Curriculum konnte entweder als Einheit über sechs Wochen durchgeführt werden oder in den Sozialkundeunterricht integriert werden. Die Einstellungen wurden mit der Ethnic Preference Scale (Westphal, 1977 aufbauend auf dem Preschool Racial Attitude Measure II) erfasst. Die Ergebnisse zeigten signifikant mehr positive Bewertungen von schwarzen Kindern in der Experimentalgruppe und signifikant mehr positive Bewertungen von weißen Kindern in der Kontrollgruppe.

Die hier kurz vorgestellten älteren Studien sind größtenteils Belege für die positive Wirkung von wissensbasierten Interventionen auf die intergruppalen Einstellungen von Schülern. Allerdings sollten diese Ergebnisse mit gewisser Vorsicht betrachtet werden, da gerade in früheren Studien häufig methodische

³ Außerdem wurden vier verschiedene Skalen zur Messung der Einstellungen verwendet, die nur gering miteinander korrelierten. Dies spricht dafür, dass Vorurteile ein komplexes Konstrukt sind, das verschiedene Facetten aufweist, welche mit unterschiedlichen Instrumenten abgebildet werden müssen (Katz, 1976).

Mängel enthalten sind, welche die Zuverlässigkeit und Generalisierbarkeit ihrer Ergebnisse stark einschränken.

Aktuellere Evaluationsstudien zu wissensbasierten Interventionen

Aktuellere Studien sollten weniger methodische Mängel aufweisen (auch wenn sie oft nicht ganz frei davon sind) und führen so zu zuverlässigeren Aussagen über die Wirksamkeit einzelner Interventionen. Im Folgenden werden zwei aktuellere Evaluationsstudien ausführlich vorgestellt.

In einer Feldstudie mit $N = 79$ Vorschulkindern aus sechs verschiedenen Vorschulen in New Hampshire untersuchten Perkins und Mebert (2005) die Wirksamkeit von multikulturellen Curricula. Die Kinder waren im Durchschnitt fünf Jahre und vier Monate alt und wurden zu drei verschiedene Gruppen mit unterschiedlichen Unterrichtsansätzen zusammengefasst: 1. Multikulturelles und kindzentriertes Curriculum ($n = 43$). 2. Multikulturelles und nicht-kindzentriertes Curriculum ($n = 20$). 3. Weder multikulturelles noch kindzentriertes Curriculum ($n = 16$). Die Kinder wurden individuell mit verschiedenen Instrumenten getestet. Dabei zeigte sich, dass die Kinder aus der ersten Gruppe mehr Gemeinsamkeiten und Unterschiede zwischen den verschiedenen Ethnien (Weiße, Schwarze, Japaner) berichten konnten als die Untersuchungsteilnehmer aus den anderen beiden Gruppen. Also verfügten die Kinder aus der ersten Gruppe über mehr ethnienspezifisches Wissen. Weiterhin konnten die Autoren feststellen, dass die Kinder aus den Gruppen eins und zwei, die ein multikulturelles Curriculum in der Vorschule besuchten, insgesamt mehr positive als negative Zuschreibungen beim Multi-Response Racial Attitude Measure (MRA) zuordneten. Allerdings fand diese positivere Einschätzung gegenüber allen drei oben genannten Zielgruppen statt, so dass sich dabei keine Unterschiede im Level der gruppenbezogenen sozialen Einstellungen zwischen den Interventionsgruppen und der Kontrollgruppe ergaben. Die hier verwendeten multikulturellen Curricula ermöglichen den Kindern also eine besser strukturierte und differenzierte Wissensbasis, besonders wenn der Unterricht sich um die Interessen und Belange der Kinder dreht. Außerdem scheinen multikulturelle Curricula dazu zu führen, dass Menschen positiver bewertet werden und allgemein in einem positiveren Licht gesehen werden. Einen Einfluss auf Vorurteile konnten die Autoren in dieser Studie allerdings nicht finden. Als Kritikpunkt an der Studie kann festgehalten werden, dass die Autoren nicht berichten, welche

Inhalte genau in den verschiedenen Curricula behandelt wurden (z.B. auf welche Ethnien im Unterricht gezielt eingegangen wurde etc.) und wie lange und intensiv die einzelnen Kinder am Unterricht teilgenommen hatten, von einer Implementationsanalyse ganz zu schweigen.

Lauren Whetstone untersuchte in ihrer Doktorarbeit (1992) die Wirksamkeit von „A World of Difference-Materialien“ zur Reduktion von Vorurteilen. Dabei handelt es sich um die amerikanische Ursprungsvariante des oben vorgestellten Programms „Eine Welt der Vielfalt“. Sie nutzte ein Prä-Post-Kontrollgruppendesign mit drei Altersgruppen von Schülern: Drittklässler ($n = 107$), Siebtklässler ($n = 123$) und Elftklässler ($n = 145$). Die Interventionsklassen aus verschiedenen Schulen in Los Angeles erhielten über zehn Wochen verteilt zehn Einheiten aus den A World of Difference-Materialien. Die etwa gleich großen Kontrollklassen erhielten nur ihr reguläres Curriculum. Die teilnehmenden und von der Anti-Defamation League trainierten Lehrer der verschiedenen Klassen konnten aus den altersgerechten A World of Difference-Materialien aus über 50 verschiedenen Lektionen wählen, welche sie in ihren zehn Einheiten nutzten. Dabei stammten die Übungen für die Grundschüler aus den fünf Bereichen: „Selbst und Ähnlichkeiten“, „Wertschätzung von Unterschieden und Gemeinsamkeiten“, „Vorurteile und Diskriminierung“, „Ethnien in Südkalifornien“ und „Eine Zukunftsperspektive“. Bei den Prä- und Postmessungen wurden verschiedene Konstrukte erhoben: Schulklima und Einstellungen gegenüber Immigranten, Stereotype, soziale Distanz, die Definition von Schlüsselbegriffen (Vorurteile, Diskriminierung, Stereotype, etc.), sowie Selbstwert. Die sozialen Einstellungen wurden in Form von Stereotypen und sozialer Distanz gegenüber fünf Zielgruppen abgefragt: Weiße, Schwarze, Asiaten, Latinos und Juden. Die Ergebnisse der Untersuchung zeigen allerdings kaum bedeutsame Interventionseffekte. So finden sich keine Interventionseffekte bei der Definition von Schlüsselbegriffen für die Dritt- und Siebtklässler. Nur die Elftklässler aus ethnisch heterogenen Schulen profitieren in diesem Bereich von der Intervention. Die Einstellungen der Kinder aus den Interventionsklassen zeigten keine größeren Veränderungen als die Einstellungen der Kinder aus den Kontrollklassen. Es fanden sich in beiden Gruppen keine Hinweise für ein interaktiveres Schulklima noch für positivere Einstellungen gegenüber Immigranten oder den Wunsch nach engeren sozialen Beziehungen zu Mitgliedern anderen Ethnien. Es fanden sich zwar mehr nicht-stereotype Antworten im Posttest als im Prätest – dies jedoch sowohl in der

Interventions- als auch in der Kontrollgruppe. Insgesamt ergibt die Evaluation, dass das A World of Difference Programm, zumindest in dieser Art der Anwendung, keine Veränderungen in den ethnischen Einstellungen, dem Wissen oder dem Selbstwert der teilnehmenden Kinder im Vergleich zur Kontrollgruppe bewirkt. Diese höchst ernüchternden Ergebnisse lassen sich wahrscheinlich auf eine zu geringe Programmintensität sowie auf die unstandardisierte Durchführung der Intervention zurückführen. Die Lehrer konnten frei wählen welche Lektion gerade in ihren Unterricht passt und oft wurden die zehn Einheiten über mehr als die zehn vorgegebenen Wochen verteilt. Zudem könnten die Ergebnisse auf einen Deckeneffekt in einigen der Einstellungsmaßen zurückzuführen sein.

In den beiden vorgestellten aktuellen Studien sind die Belege für die Wirksamkeit von wissensbasierten Interventionen nicht mehr so deutlich wie in den älteren Studien. Gerade eine Standardisierung der verwendeten Programme in allen Interventionsgruppen sowie genaue Dokumentationen über die durchgeführten Übungen wären hier wünschenswert, um Ursachen für die geringe Wirksamkeit identifizieren zu können. Da zwei Studien nur einen kleinen Einblick zum Forschungsstand zu wissensbasierten Interventionen liefern können, wird an dieser Stelle zusammenfassende Literatur in Form von Metaanalysen zu diesem Thema ausgewertet

Metaanalysen zur Wirksamkeit wissensbasierter Interventionen

Es gibt mittlerweile genügend empirische Studien zur Wirksamkeit von wissensbasierten Interventionen, um deren Ergebnisse überblicksartig oder metaanalytisch zusammenfassen zu können. Die meisten Studien kommen letztlich zu positiven Ergebnissen durch wissensbasierte Programme, allerdings bleiben die Interventionseffekte, ähnlich wie bei Kontakt-Interventionen, dabei meist eher klein. So konnten Stephan, Renfro und Stephan (2004) in einer neueren Metaanalyse über 35 Studien zu verschiedenen Multicultural Education Programmen mit unterschiedlichen Altersgruppen eine Effektstärke von $d = 0.25$ für kurzfristige Einstellungsänderungen und $d = 0.38$ für kurzfristige Verhaltensänderungen feststellen. Die langfristigen Effekte der Einstellungs- und Verhaltensänderungen sind deutlich größer und liegen bei $d = 0.80$ (Einstellung) und $d = 0.86$ (Verhalten). Eine verspätete Vergrößerung der Wirksamkeit von Interventionsprogrammen ist in der Literatur als so genannter „Sleeper-Effect“ bekannt und tritt auf, wenn

Interventionsinhalte erst nach längerer Verarbeitungszeit oder zusätzlichen Lernerfahrungen ihre Wirkung voll entfalten. Allerdings umfasst die Metaanalyse von Stephan und Kollegen nur sehr wenige Studien, die langfristige Effekte untersuchen, womit die Generalisierbarkeit dieser Effekte fragwürdig bleibt. Weiterhin ist nicht klar, welche Komponente der Programme die Wirksamkeit ausmacht. Lediglich die Kontakthypothese lässt sich nach der Metaanalyse auch im Rahmen von Multicultural Education Programmen stützen: Programme, welche Begegnungen mit Fremdgruppenmitgliedern beinhalten, zeigten signifikant höhere Verhaltensänderungen als die Programme ohne Kontakt.

In einer anderen Metaanalyse zur Wirksamkeit von Anti-Rassismustrainings und Rollenspielen zur Reduktion von Vorurteilen bei Schülern, findet McGregor (1993) ebenfalls positive Effekte des Trainings auf die Einstellungen der Kinder. Beim Anti-Rassismustraining werden Themen wie Vorurteile, Stereotype und Diskriminierung von Minderheiten in der Gesellschaft und deren historische Wurzeln sowie aktuelle Ausdrucksformen diskutiert. Somit wird auch hier auf Aufklärung gesetzt und eher auf kognitive Anteile sozialer Einstellungen fokussiert. McGregor berechnet aus sieben Studien eine Effektstärke des Anti-Rassismustrainings von $d = 0.48$ auf ethnische Vorurteile. Wobei nach einer Moderatoranalyse jüngere Kinder noch stärker von einem solchen Training profitieren können.

Zusammenfassung und Fazit zu wissensbasierten Interventionen

Insgesamt konnten in zahlreichen Studien Belege für die Wirksamkeit von wissensbasierten Interventionen auf die ethnischen Einstellungen von Schülern gesammelt werden. Dabei sind allerdings wichtige Fragen der Interventionsforschung offen geblieben oder nur ansatzweise beachtet worden: Welche Bestandteile der Interventionen und welche damit verbundenen Prozesse stehen hinter ihrer Wirksamkeit? Welche Materialien sind für welche Gruppen und Interventionsziele besonders geeignet? Auch die Frage nach der notwendigen Intensität und Dauer der Interventionsprogramme ist noch nicht ausreichend geklärt. All dies sind wichtige Faktoren, die bei der Planung und Durchführung einer wissensbasierten Intervention zur Förderung positiver gruppenbezogener Einstellungen beachtet werden müssen.

II.3.5. Förderung individueller Kompetenzen

Die Idee, individuelle Kompetenzen zur Beeinflussung sozialer Einstellungen zu fördern, beruht auf dem Gedanken, dass bestimmte interindividuelle Unterschiede, zum Beispiel aus dem Bereich sozial-kognitiver Fähigkeiten, die Entwicklung und Aufrechterhaltung sozialer Einstellungen maßgeblich mitbestimmen. Dieser Grundgedanke, dass bestimmte individuelle Voraussetzungen eine Person „anfälliger“ für die Entwicklung von Vorurteilen und diskriminierendem Verhalten machen, findet sich unter anderem bereits bei Piaget und Weil (1951) und wird von Aboud in ihrer Social-Cognitive Developmental Theory of Prejudice (1988) aufgegriffen (s. Kapitel II.2.1. Entwicklung von Vorurteilen und Stereotypen).

In einem Überblicksartikel fasst Levy (1999) die empirische Forschung zu diesem Thema zusammen. Sie beschreibt sozial-kognitive Faktoren, die einen Einfluss auf die sozialen Einstellungen von Kindern haben: So konnten Bigler und Liben (1992, 1993) zeigen, dass die Fähigkeit Personen auf verschiedenen Dimensionen gleichzeitig klassifizieren zu können mit Stereotypen und dem Erinnern von stereotyp-inkonsistenten Informationen zusammenhängt. Auch die Fähigkeit verschiedene Perspektiven einnehmen zu können bzw. die abweichende Perspektive eines Fremdgruppenmitglieds als rechtmäßig akzeptieren zu können, hängt mit verschiedenen kognitiven Vorurteilsmaßen gegenüber unterschiedlichen Ethnien zusammen (Doyle & Aboud, 1995; Black-Gutman & Hickson, 1996; Steudel, 2010). Die Fähigkeit zur differenzierten Wahrnehmung von Gruppen zeigt sich ebenfalls als Prädiktor von sozialen Einstellungen. So ist zum einen die Fähigkeit Ähnlichkeiten zwischen Gruppen wahrnehmen zu können mit der Entwicklung von Vorurteilen beziehungsweise toleranten Einstellungen verknüpft (Doyle & Aboud, 1995; Black-Gutman & Hickson, 1996), zum anderen aber auch die Fähigkeit Unterschiede innerhalb von Gruppen erkennen zu können (Doyle & Aboud, 1995; Black-Gutman & Hickson, 1996; Katz, Sohn & Zalk, 1975). Dabei scheinen diese Zusammenhänge zwischen gruppenbezogenen sozialen Einstellungen und sozial-kognitiven Fähigkeiten besonders zu kognitiven Vorurteilsmaßen, wie der Zuschreibung bestimmter Eigenschaften (wie beim MRA oder PRAM II) zu bestimmten Gruppen (Doyle & Aboud, 1995; Black-Gutman & Hickson, 1996; Steudel, 2010), oder der Informationsverarbeitung stereotyp-inkonsistenter Informationen (Bigler & Liben, 1992, 1993) zu bestehen. Weiterhin finden sich besonders Zusammenhänge zu dem sogenannten Counterbias, also einer negativeren Einstellung gegenüber der

Eigengruppe und einer positiveren Einstellung gegenüber der Fremdgruppe (Doyle & Aboud, 1995; Black-Gutman & Hickson, 1996). Dies spricht für die Wirkung einer differenzierteren Wahrnehmung von sowohl Eigen- als auch Fremdgruppenmitgliedern auf die entsprechenden sozialen Einstellungsmaße.

Wirkprozesse der Förderung individueller Kompetenzen

Festzuhalten bleibt zunächst, dass verschiedene kognitive und sozial-kognitive Mechanismen bei der Entstehung und Veränderung von sozialen Einstellungen eine Rolle spielen (s. Kapitel II.2. Die Entwicklung gruppenbezogener sozialer Einstellungen). So vermuten bereits Piaget und Weil (1951), dass für eine tolerante Einstellung gegenüber Personen aus fremden Ländern, das Kind eine Entwicklung weg von seiner egozentrischen Sichtweise und hin zu einer reziproken Perspektive durchlaufen muss. Dies halten die Autoren für zentral für eine kognitive Unvoreingenommenheit und affektive Offenheit gegenüber Fremden. Hier spielen also die Fähigkeit zur Perspektivenübernahme und die Anerkennung unterschiedlicher Perspektiven eine zentrale Rolle für positive gruppenbezogene soziale Einstellungen.

Aboud differenziert die Idee Piagets weiter aus und beschreibt eine kognitiv elaborierte und perzeptuell differenzierte Wahrnehmung als Mechanismus hinter der Beeinflussung oder Veränderung von sozialen Einstellungen (s. Kapitel II.2.2. Entwicklungstheorien). Fähigkeiten wie Racial Reconciliation (Doyle & Aboud, 1995; Aboud, 2002), multiple Klassifikationsfähigkeit (Bigler & Liben 1992, 1993), oder die Wahrnehmung von Gemeinsamkeiten zwischen Gruppen (Aboud & Fenwick, 1999) werden von ihr als Beispiele wichtiger Wirkfaktoren bei der Beeinflussung von Vorurteilen und Diskriminierung genannt.

Diese Wirkprozesse lassen sich auch den von Dovidio und Kollegen (2004) berichteten kognitiven Wirkprozessen als Mediatoren von Interventionen zur Beeinflussung sozialer Einstellungen zuordnen: 1. Soziale Kategorisierung. 2. Differenzierte Wahrnehmung der Eigen- und Fremdgruppe. 3. Erlernen von Verhaltensnormen zum Umgang mit Fremdgruppen. Diese wurden bereits an anderer Stelle ausführlich erläutert. (s. Kapitel II.3.4. Wissensbasierte Interventionen). In verschiedenen Studien werden Interventionen, die auf diesen Wirkprozessen zur Beeinflussung gruppenbezogener sozialer Einstellungen beruhen,

verwendet. Einige zentrale Arbeiten aus diesem Bereich werden im Folgenden berichtet.

Empirische Befunde zur Wirksamkeit der Förderung individueller Kompetenzen

Die Idee relevante sozial-kognitive Kompetenzen zu fördern, um eine Reduktion von negativen gruppenbezogenen sozialen Einstellungen zu erreichen, wurde seit den siebziger Jahren systematisch verfolgt. Dabei zeigten unterschiedliche Studien recht erfolgreiche Ergebnisse dieses Interventionsansatzes.

Differenzierte Wahrnehmung von Gruppen

In der Arbeitsgruppe um Phyllis Katz wurde versucht, mittels Übungen zur differenzierteren Wahrnehmung von Fremdgruppenmitgliedern (hier: Afroamerikaner), Vorurteile gegenüber dieser Gruppe zu verringern. Dahinter steht die Idee, dass eine differenzierte Wahrnehmung der Fremdgruppe den Homogenitätsbias⁴ senkt und somit die Anwendung von übergeneralisierenden Einstellungen, wie Vorurteilen und Stereotypen auf diese Fremdgruppe erschwert wird.

So konnte Katz (1973) zeigen, dass verschiedene Trainings zur besseren Differenzierungsfähigkeit von Gesichtern bei Fremdgruppenmitgliedern zu einer Verringerung von Vorurteilen führen. Die Autorin hat dies in einer Untersuchung mit $N = 96$ vorurteilsbelasteten Kindern⁵ der zweiten und sechsten Klassen getestet. Die Kinder wurden dabei in drei Interventionsgruppen eingeteilt:

1. Lernen verschiedener Namen fotografierte Fremdgruppen-Gesichtern (40 Durchgänge).
2. Gleich-Ungleich Einschätzungen von Gesichter-Paaren aus der Fremdgruppe (40 Durchgänge).
3. Gesichter anschauen ohne Benennung (Kontrollgruppe, ebenfalls 40 Durchgänge).

Die Ergebnisse zeigen, dass die Kinder aus den Gruppen 1. und 2. im Posttest ein oder zwei Tage nach der Intervention weniger Vorurteile als die Kinder

⁴ Homogenitätsbias bedeutet dass Mitglieder einer Fremdgruppe als ähnlicher wahrgenommen werden, als Mitglieder der Eigengruppe oder Mitglieder verschiedener Gruppen.

⁵ Das Kriterium um als vorurteilsbelastet zu gelten waren hohe Werte (über Median) auf zwei Vorurteilsmaßen: Dem Projective Prejudice Test (Katz, 1973) und einem soziale Distanzmaß (Katz, 1973). Dies betraf $n = 96$ von $N = 472$ Kindern, also gut 20% der Kinder aus dem Prätest.

aus der Kontrollgruppe aufwiesen. Die bessere Differenzierung von Fremdgruppenmitgliedern via Benennungstraining oder gleich-ungleich-Unterscheidungsaufgabe mit Fotos von Fremdgruppengesichtern führt zu weniger Vorurteilen im Projective Prejudice Test (Katz, 1973) und auf einem sozialen Distanzmaß. Dies ist ein, ob der Kürze der Intervention, sehr positives Ergebnis, selbst wenn die Effekte nur kurzfristig getestet wurden. Die Autorin erklärt den Erfolg ihrer Intervention folgendermaßen: Vorurteilsbehaftete Kinder haben einen Fremdgruppen-Homogenitätsbias und unterscheiden wenig zwischen Mitgliedern der Fremdgruppe. Dies zeigt sich beispielsweise auch bei der Benennung von verschiedenen Fremdgruppenmitgliedern mit dem immer gleichen Gruppenlabel. Eine Differenzierung der Wahrnehmung von Fremdgruppenmitgliedern verhindert aber unter anderem eine derartige Benennung und müsste es dementsprechend schwieriger machen übergeneralisierende Vorurteile weiterhin anzuwenden.

Diese Ergebnisse konnten Katz und Zalk (1978) in einer weiteren vergleichenden Evaluationsstudie bestätigen. Hier untersuchten sie die Wirksamkeit von vier verschiedenen Interventionen zur Reduktion von Vorurteilen und diskriminierendem Verhalten bei abermals vorurteilsbelasteten weißen Kindern⁶ ($N = 140$) der zweiten und fünften Klassen. Die vier Interventionen bestanden aus:

1. Positivem intergruppalen Kontakt

Die Kinder erhielten eine Teamaufgabe. Vier Kinder, je zwei schwarze Kinder und zwei weiße Kinder, sollten ein Puzzle zusammenlegen und dabei möglichst schnell sein. Das schnellste Team gewann einen Preis.

2. Stellvertretender Identifikation

Kinder in Dreiergruppen hörten eine Geschichte und sahen Bilder dazu. Darin ging es um ein schwarzes Fremdgruppenkind, das aus der Schule nach Hause geht und seine kranke Großmutter ins Krankenhaus begleitet.

3. Positiver Verstärkung der Farbe schwarz

Kinder sollten sich zwischen schwarzen und weißen Tieren entscheiden und erhielten immer wenn sie sich für das schwarze Tier entschieden eine Belohnung.

⁶Das Kriterium um als vorurteilsbelastet zu gelten waren hohe Werte (über Median) auf dem Katz-Zalk-Projective Prejudice Test (Zalk & Katz, 1976). Dies betraf $n = 140$ Kinder aus dem Prätest.

4. Differenziertere Wahrnehmung von Gesichtern von Fremdgruppenmitgliedern
Bilder einer schwarzen Person, die auf verschiedenen Dimensionen variierten, wurden den Kindern gezeigt. Die Kinder sollten entweder den Namen der jeweiligen Person sagen oder sich nur die Anzahl der gesehen Bilder merken.

All diese Interventionen fanden an einem Nachmittag statt und dauerten nicht länger als 15 Minuten. Im Posttest zwei Wochen nach der Intervention zeigten vor allem die Kinder aus der Wahrnehmungsdifferenzierungsgruppe positivere Intergruppeneinstellungen auf dem Katz-Zalk-Projective Prejudice Test (Zalk & Katz, 1976) und der Koslin Social Distance Scale (Koslin, Amarel & Ames, 1969). Die Interventionen wirkten sich bei den Kindern sogar auf das Verhalten (gemessen über soziale Distanz im Raum, Imitation beim Bildermalen und die Auswahl von Spiel- und Sitzplatz von entweder einem schwarzen oder einem weißen Versuchsleiter) aus. Auch die Intervention mit stellvertretender Identifikation zeigte sich als recht erfolgreich in der kurzfristigen Einstellungsänderung. Die Autoren schließen aus den Ergebnissen, dass es Kindern, die Stimuli differenzierter wahrnehmen können, schwerer fällt, diese übergeneralisiert also stereotyp zu bewerten. Weiterhin fanden Katz und Zalk, dass gerade jüngere Kinder in ihren Einstellungen von Identifikations-Geschichten profitieren (vgl. Kapitel II.3.3. Stellvertretender-Kontakt-Interventionen).

Im Follow-up nach vier Monaten zeigte sich nur noch die Wahrnehmungsdifferenzierungsintervention auf dem projektiven Vorurteilsmaß und auf dem Verhaltensmaß zur sozialen Distanz wirksam. Die Intervention der stellvertretenden Identifikation war nur noch auf einzelnen Subskalen bedeutsam. Insgesamt zeigten sich diese Effekte besonders bei den jüngeren und weniger bei den älteren Kindern. Die anderen beiden Interventionsformen, positiver intergruppaler Kontakt und positive Verstärkung der Farbe schwarz, blieben in dieser Form beim Follow-up allerdings genauso wenig effektiv wie zum Posttest. Somit bleibt als Ergebnis eine positive Wirksamkeit von Differenzierungstrainings von Fremdgruppenmitgliedern auf die sozialen Einstellungen dieser Fremdgruppe gegenüber.

Die Idee eine differenziertere Sichtweise über Fremdgruppenmitglieder zu fördern, um Vorurteile zu senken, haben Aboud und Fenwick (1999) in ihrem Artikel aufgegriffen. In zwei aktuelleren Studien konnten die Autoren zeigen, dass eine differenziertere Betrachtung von Personen zu einer Verringerung von Vorurteilen führen kann. Die erste Evaluationsstudie baut auf Erkenntnissen aus den oben

berichteten Untersuchungen zu Zusammenhängen zwischen sozial-kognitiven Kompetenzen und sozialen Einstellungen auf. Speziell die Fähigkeit zur Differenzierung von Personen innerhalb einer Fremdgruppe und ein verstärkter Fokus auf interne individuelle Eigenschaften, im Vergleich zu äußeren Gruppenzugehörigkeitsmerkmalen, hängen mit Vorurteilen zusammen und können diese verringern (Black-Gutman & Hickson, 1996; Doyle & Aboud, 1995; Katz & Zalk, 1978). Hier wurde von den Autoren ein elfwöchiges Programm evaluiert, das die individuelle Betrachtung von Personen fördern soll. Für die Studie (Aboud & Fenwick, 1999) konnte eine Stichprobe von $n = 80$ weißen ($n = 34$ Interventionsgruppe / $n = 46$ Kontrollgruppe) und $n = 46$ schwarzen Kindern ($n = 21$ Interventionsgruppe / $n = 25$ Kontrollgruppe) der fünften Klasse aus ethnisch gemischten Schulen der Mittelklasse gewonnen werden.

Das Programm basiert auf der Arbeit von Bowers und Swanson (1988) mit dem Titel „More Than Meets the Eye“. Es umfasst elf Aktivitäten, die etwa über elf Wochen verteilt stattfanden, mit ein bis zwei Sitzungen pro Woche. Die Aktivitäten beinhalten Gruppendiskussionen, dyadisches Problemlösen und individuelle Arbeit. In einer Übung sollen die Kinder beispielsweise ihr äußeres und ihr inneres Ich beschreiben und im Anschluss daran ihre Profile mit einem Freund und einem weniger gut bekannten Klassenkameraden austauschen. Häufig stellen sie dabei fest, dass sie einige Unterschiede von ihren Freunden trennen und sich unerwartete Gemeinsamkeiten mit den eher fremden Klassenkameraden finden.

Weiterhin gibt es direkte Aufklärung zum Thema Vorurteile und Stereotype sowie einen Erfahrungsaustausch über eigene Stereotype und den Umgang damit selbst Opfer von Stereotypen zu sein. Im Programm geht es allerdings nicht nur um ethnische Vorurteile und Stereotype sondern auch um gruppenbezogene soziale Einstellungen allgemein, egal auf welcher Gruppenbasis diese vorliegen (Kinder/Jugendliche, Senioren, etc.).

Die Untersuchung wurde in einem Prä-Post-Kontrollgruppendesign durchgeführt. Der Prätest fand direkt vor dem Interventionsbeginn statt. Der Posttest wurde fünf Monate später (zwei Monate nach Beendigung des Programms) durchgeführt. Dabei wurden die Anzahl und die verbale Ausdrucksweise von Unterschieden und Gemeinsamkeiten von Personen innerhalb einer Ethnie (Weiße, Schwarze, Chinesen und Natives) an Hand von Fotos beurteilt. Weiterhin wurden die Vorurteile der Kinder gegenüber Weißen, Schwarzen und Chinesen mittels des Multi-

Response Racial Attitude Measure (MRA; Doyle & Aboud, 1995) ermittelt. Dieses Maß wurde auch genutzt, um die Teilnehmer nach dem Prätest in hoch versus gering vorurteilsbelastete Kinder einteilen zu können⁷.

In der weißen Interventionsgruppe bleibt auch nach der Intervention der Homogenitätseffekt gegenüber Schwarzen bestehen, allerdings beschreiben die Kinder die Fremdgruppenmitglieder mit mehr internalen Eigenschaften als vor der Intervention im Vergleich zur Kontrollgruppe. Die Vorurteile der hoch vorurteilsbelasteten weißen Kinder sanken auf das Niveau der gering vorurteilsbelasteten Teilnehmer. Bei den gering vorurteilsbelasteten Kindern blieb das Vorurteilsniveau stabil.

In der schwarzen Interventionsgruppe fanden sich keine Interventionseffekte auf der schwarzen Homogenitätsskala. Weiterhin fanden sich keine Unterschiede zwischen den Gruppen auf der Einstellungsskala, aber auch keine Zunahme von Vorurteilen durch das Programm. Außerdem fanden sich für die schwarzen Teilnehmer auch keine Zusammenhänge zwischen Einstellungen und der Fähigkeit individuelle Unterschiede erkennen zu können.

Das Programm eignet sich also zur Reduktion von Vorurteilen bei weißen hoch vorurteilsbelasteten Kindern und verschlechtert dabei nicht die Einstellungen der gering vorurteilsbelasteten Kinder. Dabei lernen die Kinder, Paare gleicher Ethnie verstärkt mit internen Attributen zu beschreiben und zu unterscheiden. Insgesamt differenzieren sie aber Paare gleicher Ethnie nicht stärker als vor der Intervention. Der Homogenitätsbias bleibt also bestehen. Unabhängig davon ist nicht klar, welche Komponente der Intervention hier zur Vorurteilsreduzierung führte. Neben der verstärkten Beschreibung von Fremdgruppenmitgliedern mit internalen Attributen könnten genauso gut die ebenso enthaltene direkte Thematisierung von Vorurteilen und Diskriminierung wie in einem Anti-Rassismustraining den Effekt erzeugt haben.

In der zweiten Evaluationsstudie von Aboud und Fenwick (1999) sollten hoch vorurteilsbelastete Kinder durch dyadische Diskussionen mit toleranteren

⁷ Der Gesamt-Vorurteilswert wurde aus der Differenz zwischen Bias-Bewertungen (positiv weiß + negativ schwarz) und Counterbias-Bewertungen (negativ weiß + positiv schwarz) berechnet. $n = 33$ Kinder mit einem Gesamt-MRA-Wert > 0 wurden als hoch vorurteilsbehaftet eingestuft. $n = 47$ Kinder mit einem Gesamt-MRA-Wert ≤ 0 wurden als gering vorurteilsbelastet eingestuft.

Gleichaltrigen dazu gebracht werden, Fremdgruppenmitglieder differenzierter zu betrachten und somit weniger übergeneralisierende Vorurteile zu zeigen.

Dabei werden drei zentrale Argumentationslinien als effektiv angenommen: 1. Das Betonen von Ähnlichkeiten zwischen Ethnien. 2. Das Betonen von Unterschieden innerhalb von Ethnien. 3. Das Vorbringen von Beispielen, die die negative Bewertung der Fremdgruppe widerlegen. Die Stichprobe in dieser Untersuchung setzte sich aus $N = 88$ weißen Kindern (38 Jungen und 50 Mädchen) im Alter zwischen acht und elf Jahren zusammen. Die Kinder wurden zunächst mit dem Multi-Response Racial Attitude Measure (MRA, Doyle & Aboud, 1995) getestet. Dann wurden sie nach hoch und gering vorurteilsbelasteten Kindern eingeteilt⁸ und in gemischten Paaren zusammengebracht. Sie sollten nun, in Abwesenheit des Versuchsleiters, für zwei Minuten diskutieren welche Eigenschaften im MRA welcher Gruppe zugeordnet werden sollten. Diese Diskussion wurde aufgezeichnet. Zwei Minuten später betrat der Versuchsleiter den Raum und führte mit den Kindern abermals den MRA für den Posttest durch.

Die Auswertung der Daten zeigte, dass die gering vorurteilsbelasteten Kinder in der Diskussion mehr negative Äußerungen über Weiße und mehr Äußerungen zu Ähnlichkeiten zwischen Ethnien machten, als die hoch vorurteilsbelasteten Kinder. Die Posttest-MRA-Werte blieben bei den gering vorurteilsbelasteten Kindern stabil auf dem Niveau des Prätests. Bei den hoch vorurteilsbelasteten Kindern hingegen sanken die Posttest-MRA-Werte nach der Diskussion signifikant ab. Dieses Absinken hing bedeutsam mit der Anzahl bestimmter Äußerungen des toleranten Diskussionspartners zusammen: Positive Aussagen und Beispiele über Schwarze und negative Aussagen und Beispiele über Weiße sowie Hinweise auf ethnienübergreifende Ähnlichkeiten korrelierten mit dem Absinken der Vorurteile.

Fazit: Trotz offener Diskussion, also ohne die Aufgabenstellung den Anderen zu einer anderen Einstellung zu bewegen, bewirkt die Diskussion mit dem gering vorurteilsbelasteten Peer eine Verringerung der Vorurteile bei den hoch vorurteilsbelasteten Kindern. Allerdings erfolgte der Posttest direkt nach dem Prätest und der Intervention und die Messung hängt direkt mit der Intervention zusammen. Daher sollten diese Ergebnisse mit Vorsicht interpretiert und nicht unbedingt generalisiert werden. Dennoch sollte der Hinweis, dass Diskussionen über eine

⁸ Die Kinder wurden an Hand des Medians der MRA-Werte in hoch und gering vorurteilsbelastete Kinder eingeteilt.

differenzierte Betrachtung von Eigen- und Fremdgruppe Vorurteile abbauen können, ernst genommen werden. Gerade Peers können so als wertvolle Ressource genutzt werden.

Kritisch bleibt an dieser Stelle noch anzumerken, dass die positiven Effekte der Interventionen in den oben beschriebenen Studien (Katz, 1973; Katz & Zalk, 1978; Aboud & Fenwick, 1999) zur differenzierteren Wahrnehmung von Fremdgruppenmitgliedern einzig bei Kindern, die im Vergleich zu ihren Altersgenossen hoch vorurteilsbelastet waren, gewirkt haben. Kinder mit einem niedrigeren Vorurteilsniveau erfahren keine positive Beeinflussung durch die Trainings. Somit könnte der positive Effekt der Interventionen zum Teil auch auf einer Regression zur Mitte beruhen. Weiterhin könnten für den Trainingserfolg beider Studien ebenso das Vorleben positiver gruppenbezogener sozialer Einstellungen als Gruppennormen durch den Trainer oder das Peerkind verantwortlich sein. Dieser mögliche Wirkprozess wurde in den Studien jedoch nicht kontrolliert.

Multiple Klassifikationsfähigkeit

Ein weiterer Ansatz, der dazu dienen soll Personen komplexer wahrzunehmen, liegt in der Förderung der Fähigkeit zur multiplen Klassifikation. Diese Fähigkeit ermöglicht es, Personen in mehrere Kategorien gleichzeitig einzuordnen. Dies sollte dazu führen, dass Personen nicht mehr nur anhand eines einzigen Merkmals, wie zum Beispiel der Hautfarbe, bewertet werden müssen, sondern dass komplexere Personenbetrachtungen basierend auf Mehrfachkategorisierungen (vgl. Dovidio et al., 2000; Stephan & Stephan, 2001) als Bewertungsgrundlage herangezogen werden können und somit Vorurteile und Stereotype gesenkt werden.

Bigler und Liben (1992) befassten sich dazu mit der Wirksamkeit verschiedener sozial-kognitiver Interventionen auf Geschlechtsstereotype und auf die Verarbeitung stereotyp-inkonsistenter Informationen. Sie untersuchten $N = 75$ fünf bis zehn Jahre alte Kinder und teilten diese gematched nach ihren Prätestwerten in multipler Klassifikationsfähigkeit sowie ihrer Geschlechtsstereotype in vier möglichst ähnliche Gruppen ein. Die erste Gruppe erhielt ein Training in multipler Klassifikationsfähigkeit mit nicht-sozialen Stimuli (Gegenständen). Die zweite Gruppe erhielt das gleiche Training, allerdings mit sozialen Stimuli (Personen). Die dritte Gruppe lernte eine neue Regel zur Berufswahl: Demnach sei die Berufswahl eines

Menschen nicht von seinem Geschlecht abhängig, sondern von seinen Interessen und Fähigkeiten. Die vierte Gruppe schließlich erhielt normalen Unterricht zum Thema Berufe und diente somit als Kontrollgruppe. Das Training fand täglich für 25 Minuten über eine Woche statt. Der Prätest fand in der Woche vor dem Training statt und der Posttest in der Woche danach. Dabei kamen ein Multiple Classification Skill Test (Bigler & Liben, 1992), ein Test für Geschlechtsstereotype, die Gender-Stereotyped Attitudes Scale for Children (GASC, Signorella & Liben, 1985), sowie ein Gedächtnistest für stereotyp-inkonsistente Informationen zum Einsatz. Zur Auswertung wurden die Kinder nach ihrer Trainingsgruppe und ihren Ergebnissen im Multiple Classification Skill Test (MC-Test) in fünf Gruppen eingeteilt:

1. NON (Kinder die weniger als 75% der Punktzahl im MC-Test erreichten und in der Kontrollgruppe waren)
2. Class-nonsocial (Klassifikationstraining mit nicht-sozialen Stimuli und Erreichen von mindestens 75 % der Punkte im MC-Test)
3. Class-social (Klassifikationstraining mit sozialen Stimuli und Erreichen von mindestens 75 % der Punkte im MC-Test)
4. NON-rules (Regeltraining aber weniger als 75% der Punktzahl im MC-Test)
5. Class-rules (Regeltraining und Erreichen von mindestens 75 % der Punkte im MC-Test)

Als Ergebnis konnten die Gruppen nach der Anzahl der nicht-stereotypen Antworten im GASC unterschieden werden. Gruppe fünf hat am meisten nicht-stereotype Antworten und Gruppe eins am wenigsten. Weiterhin erinnerten alle Gruppen mehr gegenstereotype Informationen aus den Geschichten des Gedächtnistest als die Kontrollgruppe. Bessere Klassifizierer zeigten weniger stereotype Antworten. Das Training macht deutlich, dass das Üben von Klassifikationsfähigkeit mit sozialen Stimuli (auf denen Informationen über Geschlecht und Beruf nicht stereotyp verknüpft sind) zu geringeren Geschlechtsstereotypen führt. Dieser Zusammenhang zwischen multipler Klassifikationsfähigkeit und Geschlechtsstereotypen konnte auch korrelativ im Prätest bestätigt werden. Weiterhin scheint die Regeltechnik auch effektiv zur Senkung von Stereotypen zu sein, aber nicht zwingend notwendig dafür. Das Regeltraining wirkt dabei auch ohne die Fähigkeit zu multipler Klassifikation. Klassifikationsfähigkeit ist also nicht zwingend erforderlich um nicht-stereotype Antworten in Bezug auf geschlechtstypische Berufe zu geben. Die neue Regel

ersetzt vermutlich das alte Stereotyp bei der Zuordnung von Personen zu Berufen. Das Klassifikationstraining mit nicht-sozialen Stimuli hingegen reicht nicht aus, um Geschlechtsstereotype signifikant zu senken. Weiterhin sind die Kinder aus den Gruppen zwei bis fünf besser im Erinnern von gegenstereotypem Wissen aus Geschichten als die Kinder aus der Kontrollgruppe, die schlechter in ihrer Klassifikationsfähigkeit sind. Multiple Klassifikationsfähigkeit erlaubt also Stimuli auch gegen die vorhandenen Stereotype zu verarbeiten und zu erinnern. Mit dieser Studie konnten Bigler und Liben zeigen, dass sowohl das Training multipler Klassifikationsfähigkeiten (besonders das auf die zu verändernden Stereotype bezogene, mit sozialen Stimuli), aber auch direktes Regeltraining (in dem Kindern andere Kriterien als das Geschlecht für die Berufswahl beigebracht werden) effektiv in der Beseitigung von Geschlechtsstereotypen zur Berufswahl sind.

Allerdings sollte an dieser Stelle festgehalten werden, dass das Training multipler Klassifikationsfähigkeiten in zwei Studien von Cameron und Kollegen (Cameron, Rutland & Brown, 2007) keine positive Wirksamkeit auf Vorurteile und Verhaltensintentionen gegenüber Fremdgruppenmitgliedern (Behinderte oder Flüchtlinge) hatte (s. Kapitel II.3.3. Stellvertretender-Kontakt-Interventionen). Dies könnte darauf hinweisen, dass das Training multipler Klassifikationsfähigkeit sich verstärkt auf inhaltlich abgestimmte kognitive Einstellungsmaße, wie hier auf Geschlechtsstereotype, auswirkt, während affektive oder konative Einstellungsanteile unbeeinflusst bleiben.

Perspektivenübernahme und Empathie

Eine andere Interventionsrichtung zur Reduzierung von Vorurteilen und Diskriminierung liegt in der Förderung von Perspektivenübernahmekompetenzen und Empathie gegenüber Fremdgruppenmitgliedern. Diese wird in einer Studie von Weiner und Wright (1973) aufgegriffen. Hier versuchen die Autoren mittels Empathie oder stellvertretender Identifikation mit Opfern von Vorurteilen und Diskriminierung, soziale Einstellungen positiv zu beeinflussen. Die Autoren testeten dazu eine Intervention, in der die Teilnehmer erfahren wie es sich anfühlt, selbst Opfer von Diskriminierung zu sein. Diese von Jane Elliot an ihrer dritten Klasse 1968 erstmals durchgeführte sogenannte Blue-Eyed/Brown-Eyed-Übung teilt die Teilnehmer in zwei Gruppen – ursprünglich nach der Augenfarbe – ein. Am ersten Tag der Übung wird die eine Gruppe bevorzugt behandelt und erhält Privilegien während die andere

Gruppe diskriminiert wird. Am Folgetag wird die andere Gruppe bevorzugt und die erste Gruppe diskriminiert (vgl. „The Eye of the Storm“, 1971). Weiner und Wright (1973) testeten die Wirksamkeit dieser Methode an einer Stichprobe von $N = 62$ Kindern aus zwei dritten Klassen, die sich gleichmäßig auf die Interventions- und die Kontrollgruppe verteilten. Die Teilnehmer zeigten nach der oben beschriebenen Intervention positivere Verhaltensintentionen zu schwarzen Gleichaltrigen als die Kontrollkinder. Ebenso zeigten sie geringere Vorurteile gegenüber Schwarzen und zwar sowohl direkt nach der Intervention als auch zwei Wochen später. Die Autoren vermuten Mechanismen wie Empathie bzw. stellvertretende Identifikation mit Diskriminierungsopfern hinter der Wirksamkeit der Intervention, können diese Mediationshypothese in dieser Studie auf Grund mangelnder Datenlage jedoch nicht empirisch prüfen. Weiterhin ist ebenfalls nicht klar, ob es nicht das direkte Thematisieren von Vorurteilen und Diskriminierung ist (was als eine Art von Anti-Rassismustraining verstanden werden könnte, in dem positive intergruppale Normen vermittelt werden), was der Intervention vorausging und auch im Debriefing enthalten war, was zumindest einen Teil der Wirksamkeit der Intervention ausmacht. Somit kann die Wirksamkeit dieses Ansatzes nur als ein weiterer Hinweis für eine mögliche Wirkung von Empathie und Perspektivenübernahme interpretiert werden.

Soziale Kompetenz

Die Förderung sozialer Kompetenzen allgemein bei Kindern führt zu einer Reduktion in Anzahl und Ausmaß von Konflikten und gibt den Teilnehmern effektive Strategien zur konstruktiven Lösung von Konfliktsituationen an die Hand. Weiterhin kann ein offener, respektvoller und gerechter sozialer Umgang mit den Mitmenschen durch ein soziales Kompetenztraining gefördert werden. Somit kann auch soziale Kompetenz dazu führen, dass weniger abwertende Vorurteile oder diskriminierendes Verhalten gezeigt werden (Johnson & Johnson, 2000; Jugert, Kabak & Notz, 2006).

Johnson und Kollegen (Johnson, Johnson, Dudley & Acikgoz, 1994; Johnson, Johnson, Dudley und Magnuson, 1995) konnten in zwei Studien zeigen, dass es möglich ist, Kindern Methoden zum Umgang mit Konflikten mit Gleichaltrigen beizubringen und somit deren soziale Kompetenzen sowie ihre Fähigkeit zur Selbstregulation zu stärken.

Dazu haben die Autoren in der ersten Studie ein Peer-Mediationstraining basierend auf „Teaching Students to be Peacemakers“ (Johnson & Johnson, 1991)

an $N = 92$ Grundschulern aus vier Klassen (3.-6. Klasse) einer Schule evaluiert (Johnson et al., 1994). Die teilnehmenden Kinder wurden über sechs Wochen täglich eine halbe Stunde in Mediation und Verhandlung trainiert. Das Training erwies sich als effektiv in der Vermittlung von Mediations- und Verhandlungsfähigkeiten. Auch vier Monate nach der Intervention nutzten die Teilnehmer die gelernten Techniken, um Konflikte in der Schule selbstständig und konstruktiv zu lösen. Die Anzahl der dem Lehrpersonal vorgetragenen Konfliktsituationen hatte sich signifikant reduziert. Das Training konzentrierte sich auf die Natur des Konflikts, das Aushandeln von Vereinbarungen mit Klassenkameraden und darauf wie Konflikte zwischen Klassenkameraden mediiert werden können. Außerdem förderte es die Selbstregulation der Kinder. Das Fazit der Autoren lautet: Das Programm ist wirksam, aber der Aufwand in der Durchführung ist hoch.

In der Folgestudie (Johnson et al., 1995) kommen die Autoren an einer größeren Stichprobe von $n = 144$ Schülern der dritten bis fünften Klasse in der Interventionsgruppe und $n = 83$ Kindern in der Kontrollgruppe zu ähnlich guten Ergebnissen. Allerdings haben die Autoren die Intensität des Trainings etwas gesenkt. Insgesamt umfasste das Training in dieser Studie 9 Zeitstunden. Entweder 12-mal 45 Minuten 2 mal pro Woche, oder 18 mal 30 Minuten 3 mal pro Woche über sechs Wochen verteilt. Zusätzlich gab es Auffrischungssitzungen über das Schuljahr hinweg. Das Training zeigt sich dabei auf den folgenden Variablen effektiv: Wissen über die Verhandlungsschritte und Strategien aus dem Training, Anwendung der Strategien auf Konfliktszenarios, Anwendung der Verhandlungsstrategien in der Praxis (in verschiedenen Kontexten: Schule, Familie, Peers) sowie in einer geringeren Belastung von Lehrern und Rektoren, die deutlich weniger Konflikte zwischen den Schülern lösen mussten. Dieses Wissen und die Anwendung des Trainings blieben auch über das Schuljahr hinweg stabil.

Diese beiden Studien zeigen, dass Grundschulkinder in der Lage sind ihre sozialen Kompetenzen durch ein gezieltes Training zu steigern und somit auch lernen können, Konflikte selbstständig und konstruktiv zu lösen und respektvoll und offen mit ihrem Gegenüber umzugehen. Allerdings war der Aufwand für das Training laut den Autoren relativ hoch und die umgebenden Variablen, wie die Unterstützung des Trainings an den Schulen, waren günstig. Wenn man davon ausgeht, dass die so geförderte soziale Kompetenz der Kinder in verschiedenen Kontexten (Schule,

Familie, Peers) angewandt wird, sollte sich dies auch auf intergruppale Kontexte und somit auf gruppenbezogene soziale Einstellungen auswirken.

Zusammenfassung und Fazit zur Förderung individueller Kompetenzen

Insgesamt kann festgehalten werden, dass es unterschiedliche Ansatzpunkte bei der Förderung individueller Kompetenzen zur Reduzierung von Vorurteilen, Stereotypen und Diskriminierung gibt: Die erhöhte Differenzierung bei der Wahrnehmung von Fremdgruppenmitgliedern (Katz, 1973; Katz & Zalk, 1978; Aboud & Fenwick, 1999), die Ermöglichung einer komplexeren weniger eingeschränkten und verzerrten Informationsverarbeitung (Bigler & Liben, 1992; Aboud & Fenwick, 1999), die Förderung von Empathie und Perspektivenübernahme gegenüber Fremdgruppenmitgliedern (Weiner & Wright, 1973), oder eine allgemeine Förderung von sozialer Kompetenz bei Kindern (Johnson et al., 1994; Johnson et al., 1995). All diese Ansatzpunkte können dazu beitragen Vorurteile, Stereotype und diskriminierendes Verhalten zu senken. Ihre Wirksamkeit erfährt allerdings in der Generalisierbarkeit der bisherigen empirischen Befunde gewisse Einschränkungen. So zeigten sich einige Interventionen nur bei Kindern mit einer erhöhten Vorurteilsbelastung als wirksam (Katz, 1973; Katz & Zalk, 1978; Aboud & Fenwick, 1999). In anderen Untersuchungen konnten die eigentlichen Wirkmechanismen hinter den Interventionen nicht eindeutig bestimmt werden (Weiner & Wright, 1973 und Aboud & Fenwick, 1999 erste Studie). Weiterhin konnten manche Studien die Wirksamkeit ihrer Interventionen nur auf bestimmte Aspekte von gruppenbezogenen sozialen Einstellungen nachweisen (Bigler & Liben, 1992). Auch die Frage der Übertragung der positiven Wirkung der verschiedenen Interventionen auf weitere Fremdgruppen bleibt in den meisten Studien offen.

Ungeachtet dessen sprechen die hier vorgestellten gesammelten Ergebnisse für die Möglichkeit, mittels Förderung von individuellen sozial-kognitiven Kompetenzen bestimmte Komponenten von gruppenbezogenen sozialen Einstellungen zu reduzieren und positive Intergruppenbeziehungen zu fördern.

Im nächsten Kapitel dieser Arbeit werden die aus der Theorie ableitbaren Folgerungen für die Entwicklung und Durchführung eines multimodalen Interventionsprogramms zur Förderung von Intergruppenbeziehungen bei Kindern dargestellt. Weiterhin werden die Zielkriterien, die zur Beurteilung der Wirksamkeit

eines solchen Programms notwendig sind, sowohl normativ als auch theoretisch hergeleitet.

II.4. Schlussfolgerungen für PARTS

Nach der ausführlichen Darstellung des theoretischen Hintergrunds müssen nun die notwendigen Schlüsse für die Planung, Durchführung und Bewertung des Interventionsprogramms PARTS zur Förderung von positiven gruppenbezogenen sozialen Einstellungen gezogen werden. Dies geschieht in den nächsten beiden Abschnitten.

II.4.1. Schlussfolgerungen für die Planung und Durchführung von PARTS

Aus den oben dargestellten theoretischen Grundlagen ergeben sich verschiedene Forderungen, die an ein modernes und dem aktuellen Stand der Forschung entsprechendes Interventionsprogramm gestellt werden sollten.

Präventives Vorgehen

Aus Vorurteilsdefinitionen, sowie Studien zur Entwicklung von Vorurteilen geht hervor, dass einmal bestehende und gefestigte Einstellungen über die Zeit stabil und nur schwer veränderbar sind (e.g. Allport, 1954, Aboud, 1988, Devine, 1989). Dementsprechend sollte eine Intervention zur Verbesserung gruppenbezogener sozialer Einstellungen bereits im Kindesalter ansetzen. Dazu sollte ein Zeitfenster genutzt werden, in welchem bereits eine erhöhte Plastizität der Einstellungen vorliegt (Aboud, 1988, Nesdale 2000, Aboud & Amato, 2001, Doyle & Aboud, 1995, Raabe & Beelmann, in Druck), so dass die Intervention durch den natürlichen Entwicklungsprozess gestützt wird. Dementsprechend sollten Programme zur Prävention von negativen gruppenbezogenen sozialen Einstellungen in der mittleren Kindheit ansetzen, da dort ein Abfall der negativen gruppenbezogenen sozialen Einstellungen und ein differenzierteres Denken bei den Kindern einsetzen (Aboud, 1988; Doyle & Aboud, 1995; Raabe & Beelmann, in Druck; s. Kapitel II.2. Die Entwicklung gruppenbezogener sozialer Einstellungen).

Ganzheitlicher (multimodaler) Ansatz

Aus den Begriffsdefinitionen der verschiedenen Einstellungskomponenten (s. Kapitel Grundlagen), aber auch aus verschiedenen Entwicklungstheorien (Aboud, 1988, 2008; Nesdale, 2000, 2008; Bar-Tal & Teichman, 2005; s. Kapitel Entwicklung

gruppenbezogener sozialer Einstellungen) sowie aus der empirischen Grundlagen- und Evaluationsforschung (Dovidio et al., 2004; Bigler & Liben, 1992; Cameron et al., 2007; Howell, 1973; s. Kapitel Empirische Studien) kann geschlossen werden, dass gruppenbezogene soziale Einstellungen ein facettenreiches und uneinheitliches Konstrukt sind. Die einzelnen Einstellungskomponenten hängen zwar zusammen, jedoch nur bis zu einem gewissen Grade (Dovidio et al., 1996). Weiterhin gibt es Hinweise dafür, dass bestimmte Interventionsformen eher dazu geeignet sind bestimmte Facetten von gruppenbezogenen sozialen Einstellungen zu verändern (Dovidio et al., 2004, s. Kapitel II.3. Empirische Studien). Um eine umfassende Veränderung in diesen Einstellungen bewirken zu können, ist es daher notwendig mit der Intervention möglichst viele Bereiche gruppenbezogener sozialer Einstellungen anzusprechen. Somit ist ein multimodaler, ganzheitlicher Ansatz zur Veränderung von gruppenbezogenen sozialen Einstellungen insgesamt unumgänglich und notwendig. Dazu sollten die als änderungswirksam identifizierten und vermuteten Prozesse (Dovidio et al., 2004; Pettigrew & Tropp, 2008; s. Kapitel Entwicklung gruppenbezogener sozialer Einstellungen und Empirische Studien) aktiviert werden.

Der Schwerpunkt des Interventionsprogramms sollte in der vorliegenden Studie jedoch auf sozial-kognitiven Trainingsinhalten liegen, da die Vorurteilsentwicklung in der mittleren Kindheit (Teichmann & Bar-Tal, 2008) vornehmlich durch sozial-kognitive Prozesse beeinflusst werden.

Intensität und Dauer des Programms

Aus den empirischen Studien geht hervor, dass die Intensität und Dauer einer Intervention deren Wirksamkeit entscheidend beeinflussen (u.a. Litcher & Johnson, 1969; Litcher et al., 1973; Howell, 1973; Whetstone, 1992). Leider sind hier jedoch aus der bisherigen Forschung keine genauen Angaben zu entnehmen, welche Interventionsform wie lange und wie intensiv durchgeführt werden muss, um einen dauerhaften Effekt auf den gruppenbezogenen sozialen Einstellungen zu erzielen. Nach einer unveröffentlichten Metaanalyse von Beelmann und Heinemann (in Vorbereitung) über 62 Studien zur Prävention von Vorurteilen umfassen die durchschnittlichen Interventionen jedoch etwa 7 bis 8 Sitzungen und erzielen damit durchschnittlich kleine kurzfristige Effekte auf die gruppenbezogenen sozialen Einstellungen. Weiterhin empfehlen verschiedene Autoren, dass es für langfristige Interventionseffekte langfristige Interventionen benötigt (Duckitt, 1992b; Hill &

Augustinos, 2001). Daher sollte ein wissenschaftlich fundiertes Programm für eine dauerhafte Veränderung von Einstellungen ebenfalls mindestens 8 Sitzungen umfassen, besser mehr.

Standardisierung und Strukturierung

Um Aussagen über die Wirksamkeit des Programms machen zu können, muss sicher gestellt sein, dass in allen teilnehmenden Schulklassen das gleiche Programm durchgeführt wird (Whetstone, 1992; Lessing & Clarke, 1976). Dies kann über ein standardisiertes Interventionsmanual, standardisierte Interventionsmaterialien sowie über einheitlich ausgebildete und untereinander vernetzte Trainer sichergestellt werden. Weiterhin erweist sich ein hochstrukturiertes Vorgehen bei Interventionen im Kinderbereich als überlegen im Vergleich zu unstrukturierten Interventionen, was vermutlich der besseren Umsetzung und leichteren Implementation solcher Programm zu verdanken ist (vgl. Beelmann & Raabe, 2007).

Bebildertes entwicklungsangemessenes Interventionsmaterial

Basierend auf den Ergebnissen empirischer Untersuchungen zeigt sich, dass besonders bebildertes Material geeignet ist, um gruppenbezogene soziale Einstellungen zu beeinflussen (Litcher & Johnson, 1969; Litcher et al., 1973; Cameron et al., 2006a, 2006b, 2007, s. Kapitel Empirische Studien). Dies ermöglicht eine konkretere und kindgerechte Auseinandersetzung mit interkulturellen Themen.

Des weiteren ist es notwendig die Interventionsinhalte dem Fähigkeitsniveau der Teilnehmer anzupassen, um eine mögliche Überforderung oder Unterforderung zu vermeiden. Dazu sollte den Kindern nur die notwendige Hilfestellung gegeben werden, um sich von ihrem Fähigkeitslevel ins nächst höhere Level zu entwickeln (vgl. Zone der proximalen Entwicklung, Vygotsky, 1978), und den Kinder so optimale Lernerfolge zu ermöglichen.

Kindzentriertheit

Interventionsprogramme sollten die aktuelle Situation der Kinder, ihre Wünsche und Bedürfnisse sowie ihre Interessen berücksichtigen (Perkins & Mebert, 2005). Somit kann ein Bezug zwischen den Inhalten des Programms und dem Lebensalltag der Kinder hergestellt werden. Dies führt zu einer intensiveren

Verarbeitung und Nutzung der Inhalte und trägt zur Generalisierung des Erlernten auf andere soziale Kontexte bei.

Sozialer Kontext

Bei der Planung und Durchführung von Vorurteilspräventionen sollte der soziale Kontext, in dem die Kinder leben, beachtet werden. Besonders die vorhandenen sozialen Gruppen, die Beziehungen zwischen diesen Gruppen und die bereits existierenden gruppenbezogenen sozialen Einstellungen müssen für eine erfolgreiche Intervention berücksichtigt werden (Bar-Tal & Teichman, 2005; s. Kapitel Entwicklung gruppenbezogener sozialer Einstellungen).

Vorbildfunktion

Die Trainer des Programms sollten sich so verhalten und einbringen, dass sie als Vorbilder dienen können und die Kinder positive intergruppale Verhaltensnormen wie Toleranz, Akzeptanz und Respekt für den Umgang mit Eigen- und Fremdgruppenmitgliedern lernen können (Johnston & Hewstone, 1990; Nesdale et al., 2005; Rutland et al., 2005; Wittenbrink & Henly, 1996; Devine, 1995; s. Kapitel Empirische Studien).

II.4.2. Schlussfolgerungen für Zielkriterien von PARTS

Aus den oben berichteten theoretischen Grundlagen und empirischen Studien lassen sich Standards für Zielkriterien zur Beurteilung von PARTS ableiten.

Theoretische Fundierung

Ein wichtiges Qualitätsmerkmal eines Interventionsprogramms ist dessen theoretische Fundierung. Es sollte sichergestellt sein, dass die Interventionsplanung und -durchführung anhand des aktuellen Forschungsstandes erfolgt.

Implementation

Neben der Evaluation der Wirksamkeit des Programms sollte auch dessen Umsetzbarkeit gewährleistet sein. Da PARTS als universelles Präventionsprogramm konzipiert ist, sollte sichergestellt sein, dass dieses von geschulten Trainern bei gleich bleibender Qualität an unterschiedlichen Standorten durchgeführt werden kann und dabei die breite Masse der Schüler anspricht.

Proximale Wirksamkeit

Als Voraussetzung dafür, dass sich PARTS entsprechend der Interventionsmethoden wissensbasierte Intervention, Stellvertretender-Kontakt-Intervention und Förderung sozial-kognitiver Kompetenzen auf die Entwicklung von gruppenbezogenen sozialen Einstellungen auswirkt, muss sichergestellt sein, dass diese Methoden bei dem Programm wirkungsvoll eingesetzt werden. Um dies prüfen zu können, sollten proximale Effektmaße bezogen auf die Interventionsmethoden in der Evaluationsstudie verwendet werden. Dementsprechend sollte einer Wirksamkeitsevaluation in Bezug auf die gruppenbezogenen sozialen Einstellungen eine Wirksamkeitsevaluation bezüglich dieser proximalen Effektmaße vorausgehen.

Distale Wirksamkeit

Verminderung negativer gruppenbezogener sozialer Einstellungen

Ziel von PARTS ist eine Verminderung negativer gruppenbezogener sozialer Einstellungen. Diese kann auf unterschiedlichen Wegen erfolgen, wobei beide Vorgänge im Sinne einer Zielerreichung interpretiert werden:

Da negative gruppenbezogene soziale Einstellungen in der Kindheit zwar auftreten, aber in ihrem Ausmaß noch relativ gering sind (Nesdale, 2004) und da in

der mittleren Kindheit ein natürlicher Abfall, oder zumindest eine verstärkte Differenzierung dieser Einstellungen stattfindet (Aboud, 1988; Doyle & Aboud, 1995; Nesdale 2000; Raabe & Beelmann, in Druck, s. Kapitel II.2. Die Entwicklung gruppenbezogener sozialer Einstellungen), kann PARTS dann als erfolgreich bezeichnet werden, wenn es diesen Abfall beschleunigt oder verstärkt.

Ebenso kann es jedoch als Erfolg von PARTS gewertet werden, wenn das Programm den normalerweise in der weiteren Entwicklung auftretenden Wiederanstieg (Nesdale, 2000) in den negativen gruppenbezogenen sozialen Einstellungen hemmt.

Ein Erfolg von PARTS würde sich dementsprechend einerseits in einem stärkeren Abfall der negativen sozialen Einstellungen in der Interventionsgruppe im Vergleich zur Kontrollgruppe und andererseits in einer konstant gehaltenen negativen gruppenbezogenen sozialen Einstellung in der Interventionsgruppe im Vergleich zu wiederansteigenden negativen gruppenbezogenen sozialen Einstellungen der Kontrollgruppe zeigen.

Förderung von interpersonalen sozialen Einstellungen und sozialer Kompetenz

Ein weiteres Ziel von PARTS ist es, interpersonale soziale Einstellungen und soziale Kompetenz zu fördern. Dieses Ziel kann dann als erreicht gelten, wenn in der Interventionsgruppe im Vergleich zur Kontrollgruppe ein stärkerer Anstieg auf den interpersonalen sozialen Einstellungen und auf der sozialen Kompetenz über die Zeit stattfindet, bzw. ein stärkerer Abfall in der Intoleranz und des sozialen Problemverhaltens der Interventionsteilnehmer im Vergleich zur Kontrollgruppe verzeichnet werden kann.

Differentielle Evaluation:

Betrachtung spezifischer Einstellungskomponenten

Da einzelne Interventionsteile des multimodalen Programms verstärkt auf bestimmte Einstellungskomponenten wirken (Dovidio et al., 2004), müssen für eine differenzierte Evaluation von PARTS diese verschiedenen Einstellungskomponenten getrennt voneinander ausgewertet werden (affektive, kognitive und konative Komponente).

Risikogruppen

In der Literatur ist es gängige Vorgehensweise als Zielgruppe einer Intervention die Kinder auszuwählen, welche ein hohes Ausmaß an negativen gruppenbezogenen sozialen Einstellungen aufweisen (Katz, 1973; Katz & Zalk, 1978; Aboud & Fenwick, 1999). Exploratives Ziel dieser Studie ist es, festzustellen, ob sich das vorliegende Interventionsprogramm auf eine solche Risikogruppe in anderer Weise auswirkt, als auf die Gesamtheit der untersuchten Personen.

Generalisierung der Effekte

Nach Möglichkeit sollten durch das multimodale Programm bei den Kindern die Voraussetzungen gelegt werden, um gegenüber verschiedenen Fremdgruppen eine positive tolerante Einstellung zu entwickeln. Der Effekt sollte also nicht nur gegenüber einer spezifischen Gruppe (z.B. russischen Kindern) zu finden sein, sondern auch gegenüber anderen Fremdgruppen, die nicht explizit im Training behandelt wurden, aber dennoch in den Einstellungsmaßen mit erfasst werden. Die Trainingseffekte sollten sich also auf andere als die Zielfremdgruppe generalisieren. Eine generelle Toleranz soll gefördert werden.

Allgemeine Wirksamkeit

PARTS ist als universelles Präventionsprogramm geplant. Daher sollte es seine Wirksamkeit bei möglichst allen deutschen Kindern der dritten Klassen aus dem ländlichen Thüringer Raum entfalten. Die Wirksamkeit des Programms sollte also unabhängig von möglichen differentiellen Einflussfaktoren wie dem Geschlecht, der Intelligenz oder dem Alter der Teilnehmer gesichert sein.

Zeitliche Perspektive

Ziel eines Präventionsprogramms gegen negative gruppenbezogene soziale Einstellungen muss eine langfristige Einstellungs- und Verhaltensänderung sein. Um sowohl kurzfristige als auch langfristige oder verzögerte Effekte des Programms aufdecken und absichern zu können, ist somit ein längsschnittliches Untersuchungsdesign mit nachfolgenden Follow-up-Messungen unumgänglich.

III. Methoden

Nach den ausführlichen Betrachtungen der bisherigen Forschung zur Entwicklung von negativen gruppenbezogenen sozialen Einstellungen und zu entsprechenden Präventionsansätzen, werden im folgenden Kapitel die notwendigen methodischen Voraussetzungen für eine erfolgreiche Evaluationsstudie, sowie deren Umsetzung im Rahmen von PARTS dargestellt.

Zunächst werden das Interventionsprogramm PARTS und seine konzeptionellen Grundlagen vorgestellt. Dann wird auf die Inhalte des Programms, den Ablauf und die Durchführung der Intervention, sowie auf die Umsetzung der einzelnen Interventionsstrategien im Rahmen von PARTS eingegangen. Im Anschluss daran werden das Design der Untersuchung sowie die Untersuchungstichprobe beschrieben. Es wird auf die Durchführung der Untersuchung zu den drei hier zu Grunde liegenden Messzeitpunkten eingegangen, und anschließend werden die zum Einsatz gekommenen Messinstrumente zur Operationalisierung der unterschiedlichen Erfolgskriterien von PARTS vorgestellt.

III.1. *Konzeptionelle Grundlagen von PARTS*

Wie bereits im vorangegangenen Kapitel näher erläutert (s. Kapitel II.4. Schlussfolgerungen für PARTS), verfolgt PARTS ein *präventives Interventionskonzept*, um zu verhindern, dass sich negative gruppenbezogene soziale Einstellungen bereits im Kindesalter etablieren und festigen. PARTS ist zudem als *universelles Präventionsprogramm* konzipiert. Es wurde also nicht für eine Risikogruppe mit spezifischer Indikation entwickelt, sondern eignet sich zur Anwendung bei der Gesamtpopulation achttjähriger deutscher Kinder⁹. Diese Interventionsstrategie hat den Vorteil, eine möglichst breite Gruppe von Adressaten zu erreichen und damit auch „Risikokinder“, ohne das Problem einer möglichen Stigmatisierung, am Programm teilnehmen zu lassen (vgl. Offord, 2000).

⁹ Dabei muss das Programm je nach sozialem Kontext, in dem es angewendet werden soll, Anpassungen z.B. bezüglich der Zielfremdgruppen erfahren.

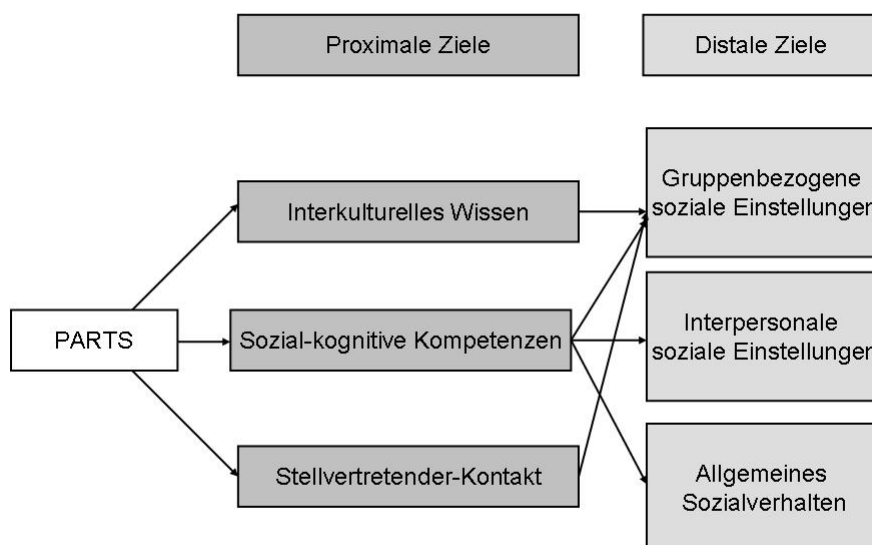
Das Interventionsprogramm PARTS wurde auf Basis der bisherigen Forschung zur Entwicklung von gruppenbezogenen sozialen Einstellungen und zu Interventionsstrategien gegen negative gruppenbezogene soziale Einstellungen entwickelt (s. Kapitel II.4. Schlussfolgerungen für PARTS). Dementsprechend verfolgt es einen multimodalen, hoch strukturierten Ansatz, um eine umfassende Einflussnahme auf möglichst viele Facetten gruppenbezogener sozialer Einstellungen zu gewährleisten, und dabei möglichst effektiv wirksam zu werden (Beelmann, 2006).

Es basiert auf drei erfolgreichen Strategien zur Reduktion von Vorurteilen, Stereotypen und Diskriminierung: Stellvertretender-Kontakt-Interventionen, wissensbasierte Interventionen und die Förderung sozial-kognitiver Kompetenzen. Da entsprechend aktueller Theorien zur Entwicklung sozialer Einstellungen davon auszugehen ist, dass in der mittleren Kindheit vor allem sozial-kognitive Prozesse auf die negativen gruppenbezogenen sozialen Einstellungen wirken, verfügt PARTS über einen sozial-kognitiven Schwerpunkt.

In der folgenden Abbildung 02 ist das Wirksamkeitsmodell von PARTS zusammengefasst dargestellt.

Abbildung 02

PARTS-Wirksamkeitsmodell



III.2. *Ablauf und Durchführung der Intervention*

Um eine möglichst wirksame Intervention zu entwickeln, wurde zum Teil auf bereits bestehende und positiv evaluierte Interventionsmaterialien zurückgegriffen, die für die hier zu erreichende Zielgruppe und den speziellen sozialen Kontext an Thüringer Grundschulen angepasst wurden. Zum Teil wurden aber auch völlig neue Übungen und Materialien erstellt, um die Zielsetzungen von PARTS zu verfolgen.

Im Folgenden werden der Ablauf und die Rahmenbedingungen des Programms, sowie die einzelnen Interventionsstrategien, wie sie bei PARTS Anwendung fanden, kurz beschrieben.

III.2.1. Gesamttablauf

PARTS beinhaltet 15 Unterrichtseinheiten, wobei die erste und die letzte Einheit der Eröffnung beziehungsweise dem Abschluss des Programms dienen. Die übrigen Einheiten beinhalten die eigentliche Intervention.

Die erste Einheit dient der Kontaktaufnahme und dem gegenseitigen Kennenlernen von Kindern und Trainern. Hier werden einige grundlegende Regeln zum sozialen Umgang erklärt, der Ablauf und das Thema des Programms vorgestellt, sowie Geschichten der Kinder und Trainer zu ihren Namen erzählt. In der letzten Einheit von PARTS werden gemeinsam anhand der PARTS-Wandzeitung die einzelnen Lektionen rekapituliert. Außerdem geben die Kinder Feedback, welche Lektionen ihnen am Besten und welche ihnen weniger gut gefallen haben. In den übrigen 13 Einheiten von PARTS findet die inhaltliche Intervention statt. Hier werden die oben beschriebenen Elemente aus den drei Bereichen Stellvertretender-Kontakt, wissensbasierte Intervention und Förderung sozial-kognitiver Kompetenzen angewendet. Der Gesamttablauf von PARTS ist Tabelle 03 zu entnehmen. Eine komplette Ausgabe des PARTS Manuals und Materials findet sich im Anhang A auf der beiliegenden DVD.

Tabelle 03

Gesamtablauf von PARTS

Lektion	Thema	Inhalt
1	Einführung in das Programm	Kennenlernen der Trainer und Kinder, Vermittlung der Programmidee, Etablierung von Regeln und Abläufen mithilfe von einleitenden Spielen
2	Wissensbasierte Intervention I	Kennenlernen der Lebensweise verschiedener Familien aus unterschiedlichen Ländern
3	Stellvertretender-Kontakt I & sozial-kognitive Kompetenz I	Lesen und Diskutieren der Geschichte „Das magische Puzzle“ und Übung der Fähigkeit zur multiplen Klassifikation der sozialen Gruppenzugehörigkeit der Kinder
4	Wissensbasierte Intervention II	Spiel zur Erkundung verschiedener Regionen und Kulturen der Erde
5	Wissensbasierte Intervention III	Fortsetzung des Spiels: Ursachen kultureller Unterschiede; Gemeinsamkeiten zwischen Kulturen
6	Sozial-kognitive Kompetenz II	Erarbeitung eines Schemas zur Lösung von alltäglichen sozialen Problemen anhand einer Fabel
7	Sozial-kognitive Kompetenz III	Üben des zuvor gelernten Problemlöseschemas sowie der Perspektivenübernahmefähigkeit
8	Sozial-kognitive Kompetenz IV & Stellvertretender-Kontakt IIa	Üben von Perspektivenübernahme anhand von Märchen aus der Sicht des Bösewichts; Lesen und Diskutieren des ersten Teils der Geschichte „Der Kinderwagenwettbewerb“
9	Stellvertretender-Kontakt IIb & sozial-kognitive Kompetenz V	Lesen und Diskutieren des zweiten Teils der Geschichte; Üben der Fähigkeit zur multiplen Klassifikation durch das Explorieren von Gemeinsamkeiten zwischen Kindern

10	Wissensbasierte Intervention IV	Übungen und Spiele zu Kontakt zwischen Mitgliedern unterschiedlicher Kulturen
11	Sozial-kognitive Kompetenz VI	Übungen zur Perspektivenübernahme und multiplen Klassifikationsfähigkeit mittels Rollen- und Bewegungsspielen
12	Stellvertretender-Kontakt IIIa & sozial-kognitive Kompetenz VII	Lesen und Diskutieren des ersten Teils der Geschichte „Top-Biker“; Übung zur multiplen Klassifikationsfähigkeit
13	Stellvertretender-Kontakt IIIb & sozial-kognitive Kompetenz VIII	Lesen und Diskutieren des zweiten Teils der Geschichte; Üben des Schemas zur sozialen Problemlösung
14	Wissensbasierte Intervention V	Spiel zur Wahrnehmung von Personen und Hinweise auf Wahrnehmungsverzerrungen
15	Abschluss des Programms	Abschluss, Rückblick und Feedback

Anmerkungen. Jede Lektion ist auf eine Unterrichtsstunde mit 45 Minuten ausgelegt. Die Lektionen 1 und 2 sowie 4 und 5 wurden als Doppelstunde durchgeführt.

III.2.2. Rahmenbedingungen der Programmdurchführung

Das Präventionsprogramm PARTS wurde im Herbst 2007 in zwölf dritten Klassen aus 12 verschiedenen Grundschulen des Schullandkreises Gotha/Bad Langensalza in Thüringen durchgeführt. Die Programmdauer betrug neun Wochen. In jeder Woche wurden zwei Einheiten zu zwei Terminen durchgeführt¹⁰. In der ersten Woche wurden jedoch drei Einheiten durchgeführt, weil die ersten beiden Lektionen als Doppelstunde zu einem Termin zusammengelegt worden waren. Dies geschah, um in der ersten Sitzung genug Zeit zu haben, einen guten Kontakt zwischen den Trainern und der jeweiligen Klasse aufzubauen, und zudem einige erste Inhalte zu bearbeiten. Die ersten drei Einheiten fanden in der Woche vom 08.-12.10.2007 und die letzte Einheit in der Woche vom 03.-07.12.2007 statt. In den zwei Wochen nach den ersten drei Einheiten fand keine Intervention statt, weil die Thüringer Herbstferien in diesen Zeitraum fielen. Die Kinder erhielten allerdings am Ende der dritten Einheit zwei kleine Ferienaufgaben, damit sie sich auch in der Ferienzeit mit Programminhalten beschäftigen¹¹.

PARTS wurde von festen Trainerteams in den Klassen umgesetzt. Jeweils zwei Trainer besuchten über die gesamte Intervention hinweg gemeinsam die gleichen Klassen. Die Trainer waren zwei wissenschaftliche Mitarbeiter (Diplom-Psychologen) und sechs sehr gut geschulte und auch an der Entwicklung des Programms beteiligte Hilfskräfte (Psychologiestudenten). Alle Trainer wurden speziell für die Umsetzung von PARTS geschult. Dabei wurde besonders auf die einheitliche und standardisierte Umsetzung der Übungen aus dem Interventionsmanual geachtet. Regelmäßige Supervisionstreffen zwischen den Trainerteams gewährleisteten diese Standardisierung.

Ablauf der einzelnen Lektionen

Die Lektionen von PARTS folgten mit Ausnahme der ersten Lektion einem einheitlichen Ablauf: Die Trainer besuchten die jeweilige Schule und Klasse und begannen nach der Begrüßung mit dem *Eröffnungsritual* die jeweilige PARTS-

¹⁰ Abgesehen von der Grundschule UF. Dort wurde auf Grund der Unterrichtsplanung, nach der die Gesamtklasse nur zu bestimmten Zeiten gemeinsamen Unterricht hatte, immer eine Doppelstunde pro Woche, statt zwei Einzeleinheiten durchgeführt.

¹¹ Die Kinder sollten einen Verwandten oder Freund mit dem „Du bist...-Arbeitsblatt“ interviewen und einen Reisereport über ihren letzten Urlaub im Ausland verfassen.

Stunde. Als Eröffnungsritual diente das Spiel „Alle in eine Reihe“, welches außerdem zur Förderung der multiplen Klassifikationsfähigkeit beitragen sollte. Hier hatten die Kinder die Aufgabe, möglichst selbst organisiert eine Rangreihe in Bezug auf ein bestimmtes Merkmal zu bilden. So sollten die Kinder sich zu Beginn der zweiten Lektion zum Beispiel anhand ihrer Körpergröße in einer Reihe aufstellen. Mit jeder weiteren Lektion sollten sich die Kinder nach einer anderen Dimension aufstellen. Wenn die Kinder dabei Schwierigkeiten hatten, griffen die Trainer unterstützend ein. Schließlich wurde überprüft, ob die von den Kindern gebildete Reihe korrekt war. Auch die Positionen einzelner Kinder wurden hervorgehoben – gerade im Vergleich zu den anderen Rangreihen, die es in vorangegangenen PARTS-Einheiten gegeben hatte. Somit diente das Eröffnungsritual nicht nur der Einstimmung auf die PARTS-Einheit, sondern ebenfalls der Förderung der multiplen Klassifikationsfähigkeit der Kinder.

Nach der Auswertung des Eröffnungsrituals nahmen die Kinder wieder ihre Plätze ein, und die Trainer begannen mit den Inhalten der entsprechenden Lektion. Zum Ende jeder Einheit wurde von ein oder zwei Kindern ein Symbol für diese Lektion an die PARTS-Wandzeitung geklebt. Diese Wandzeitung wurde in der ersten Lektion eingeführt und in der Klasse aufgehängt. Neben den Symbolen zu jeder Stunde wurden darauf die Termine für die PARTS-Einheiten vermerkt, damit die Kinder immer wussten, wann die nächste Einheit stattfand. Die Wandzeitung hatte weiterhin den Zweck, dass die Kinder auch zwischen den PARTS-Einheiten an die Lektionen erinnert wurden und sich erinnern konnten, was in den letzten Sitzungen gemacht wurde.

Nach dem Anbringen des Symbols an die Wandzeitung, wurde die PARTS-Einheit mit einem Abschlussritual beendet. Hier wurde ein kurzes Bewegungsspiel zur Auflockerung und Motivation der Kinder als Abschlussritual in die Einheiten integriert. Allerdings wurde bei der Durchführung von PARTS das Bewegungsspiel teilweise auf Grund von Zeitmangel und zu Gunsten der inhaltlichen Interventionen ausgelassen.

Während der PARTS-Stunden war immer eine Lehrkraft, meist die Klassenlehrerin, anwesend. Diese hatte die Aufgabe während der Stunde die Teilnahmebereitschaft der Kinder an den PARTS-Einheiten zu bewerten. Dazu wurden zu jeder Einheit standardisierte Implementationsbögen an die Lehrkraft

ausgegeben (s. Kapitel III.6. Messinstrumente). Die gleichen Bögen füllte nach jeder Sitzung auch einer der Trainer aus.

III.2.3. Umsetzung der Interventionsstrategien

Stellvertretender-Kontakt-Intervention

Die Idee und die Wirkprozesse von Stellvertretender-Kontakt-Interventionen sind bereits im Kapitel II.3. Empirische Studien ausführlich dargestellt worden. Die Stellvertretender-Kontakt-Intervention hat bei PARTS Eingang gefunden, da sie zum einen als sehr effizient in der Einstellungsveränderung beschrieben wird und zum anderen auch in sehr homogenen Kontexten, wie sie in Thüringen vorliegen, angewendet werden kann. Im Rahmen von PARTS wurden drei Geschichten mit stellvertretendem Kontakt gemeinsam mit den teilnehmenden Klassen bearbeitet. Sie basieren auf den Materialien, die bereits von der Arbeitsgruppe um Lindsey Cameron (Cameron & Rutland, 2006; Cameron, Rutland, Brown & Douch 2006; Cameron, Rutland & Brown 2007) erfolgreich angewendet wurden. Diese Geschichten wurden aus dem Englischen ins Deutsche übersetzt. Dabei wurden die Texte auf die Lesefähigkeit von Kindern der dritten Klasse abgestimmt. Die Fremdgruppe, mit welcher der stellvertretende Kontakt hergestellt werden sollte, wurde an den sozialen Kontext in Thüringen angepasst. Somit handeln die Geschichten von Freundschaften zwischen deutschen und russischen Kindern (statt von englischen und Flüchtlingskindern). Es wurden russische Kinder ausgewählt, da in einer Pilotstudie zur Wirksamkeit von Stellvertretender-Kontakt-Interventionen im deutschen Sprachraum, die ebenfalls in Thüringen im Rahmen dieses Projekts stattfand, sich besonders gegenüber russischen Kindern negative gruppenbezogene soziale Einstellungen fanden (Küchel, 2008). Es wurden sowohl die Namen der Charaktere in den Geschichten, als auch die größtenteils farbigen Illustrationen, die sich auf jeder Seite der Texte finden, an die beteiligten Gruppen (deutsche und russische Kinder) angepasst.

Die Kontaktvermittlung in den Geschichten wurde entsprechend des Intergruppenmodells (Hewstone & Brown, 1986) gestaltet, bei dem die Gruppenzugehörigkeit der Charaktere und deren Typikalität für ihre Gruppe salient gemacht werden. Dieses Modell hat sich in den Studien von Cameron und Kollegen als besonders erfolgreich in seiner Wirksamkeit auf gruppenbezogene soziale

Einstellungen erwiesen (s. Kapitel II.3. Empirische Studien). Dementsprechend wurden die Gruppenzugehörigkeit, die Typikalität der Charaktere und die Freundschaft zwischen den deutschen und russischen Charakteren sowohl im Text der Geschichten, als auch in den Diskussionen im Anschluss an das gemeinsame Lesen besonders betont.

Die drei ausgewählten Geschichten heißen: „Das magische Puzzle“, „Der Kinderwagen-Wettbewerb“ und „Top-Biker“. Das magische Puzzle umfasst 33 Seiten und handelt von deutschen und russischen Kindern, die gemeinsam eine Phantasiegeschichte erleben und in ein mittelalterliches Abenteuer verstrickt werden. Der Kinderwagen-Wettbewerb umfasst 34 Seiten und handelt von einer Situation in der Schule, in der gut befreundete deutsche und russische Kinder die Führung von zwei Gruppen übernehmen, die in einem Wettbewerb versuchen, den besten Kinderwagen zu entwerfen. Bei der Geschichte Top-Biker schließlich geht es in 34 Seiten um eine Gruppe von Freunden (deutsche und russische Kinder), die einen anderen Jungen nach einem Fahrradunfall retten.

Diese drei Geschichten wurden in der Klasse abwechselnd von den Kindern und den Trainern vorgelesen. Dabei wurde „Das magische Puzzle“ komplett in einer Sitzung gelesen, während die anderen beiden Geschichten, auf zwei aufeinander folgende Sitzungen verteilt, gelesen wurden. Bei der Auswahl der lesenden Schüler wurde auf die Vorlesefähigkeiten der Kinder Rücksicht genommen. Um die Sinnentnahme für die anderen Schüler jederzeit zu gewährleisten, wurde bei selten auftretenden Schwierigkeiten beim Vorlesen der entsprechende Abschnitt von einem der Trainer noch einmal vorgelesen. Im Anschluss an das Lesen erfolgte jeweils eine kurze Diskussion über das gerade Gelesene. Hier wurden von den Trainern unterschiedliche Fragen zu den Geschichten gestellt, um zum einen die Salienz der Gruppenzugehörigkeit und die Typikalität der Charaktere sowie deren Freundschaft nochmals hervorzuheben (z.B. „Wessen Familie ist aus Russland nach Deutschland gekommen?“ oder „Wer sind die besten Freunde in der Geschichte?“), und zum anderen einen Bezug der Geschichten zum Leben der teilnehmenden Kinder herzustellen (z.B. „Wer von Euch hat auch einen kleinen Bruder?“ oder „Wer von Euch hat auch einen Hund?“).

Die Kinder erhielten auch eine Kopie der Geschichten zur Aufbewahrung in ihren PARTS-Mappen¹². Somit konnten die Kinder jederzeit die Geschichten noch einmal nachlesen. Insgesamt wurden über fünf Einheiten verteilt drei Stellvertretender-Kontakt-Geschichten gelesen und besprochen, was einer Gesamtinterventionsdauer von 85 Minuten entsprach. Dabei wurden diese Einheiten immer mit Übungen zur Förderung von sozial-kognitiven Kompetenzen ergänzt. Die Aufteilung der Geschichten auf die einzelnen PARTS-Lektionen ist in Tabelle 04 dargestellt:

Tabelle 04

Stellvertretender-Kontakt-Intervention bei PARTS

Kontaktgeschichte	Inhalt	Ziel- fremdgruppe	Lektion	Dauer (Min.)
Das magische Puzzle	Befreundete deutsche und russische Kinder erleben ein mittelalterliches Abenteuer.	Russland	3	20
Der Kinderwagen-Wettbewerb Teil 1	Befreundete deutsche und russische Kinder leiten zwei	Russland	8	15
Der Kinderwagen-Wettbewerb Teil 2	Gruppen in einem Wettbewerb, den besten Kinderwagen zu entwerfen.		9	15
Top-Biker Teil 1	Befreundete deutsche und russische Kinder retten einen	Russland	12	15
Top-Biker Teil 2	anderen Jungen nach einem Fahrradunfall.		13	20

Anmerkungen. Alle Stellvertretender-Kontakt-Geschichten wurden im Intergruppenmodell angewendet.

¹² Die Kinder erhielten in der ersten PARTS-Einheit unter anderem eine eigene Heftmappe zum Abheften und Aufbewahren von Materialien, die sie im Laufe des Trainings erhalten sollten.

Wissensbasierte Intervention

Die Idee wissensbasierter Interventionen und deren Wirkprozesse sind ausführlich im Kapitel II.3. Empirische Studien beschrieben. Die für PARTS entwickelte wissensbasierte Intervention ähnelt konzeptuell dem Programm „Eine Welt der Vielfalt“ (Bertelsmann Stiftung, 2004). Sie beinhaltet fünf Lektionen, die sich anders als in den übrigen Einheiten, komplett und ausschließlich mit wissensbasierten Inhalten befassen. In diesen Lektionen wurde allerdings nicht speziell auf eine bestimmte Fremdgruppe eingegangen, sondern es wurde ein allgemeiner Ansatz mit wichtigen Erkenntnissen für eine tolerante Einstellung insgesamt verfolgt. Es wurden also in den Übungen und Beispielen Personen aus vielen verschiedenen Ländern und Kulturen vorgestellt, ohne speziell auf eine Gruppe gezielter einzugehen. Dies entspricht dem Anspruch von PARTS, eine generelle Toleranz und positive gruppenbezogene Einstellungen zu fördern (s. Kapitel II.4. Schlussfolgerungen für PARTS). Die einzelnen Übungen in den Wissenslektionen wurden zum Teil aus bestehenden Programmen und Materialien entnommen und zum Teil eigens für PARTS entwickelt.

Die Themen dieser fünf Lektionen lassen sich wie folgt zusammenfassen: 1. Wie leben die Menschen dieser Welt? 2. Die Welt-Schatzsuche 3. Gemeinsamkeiten zwischen Kulturen 4. Kontakt zwischen Kulturen 5. Jeder Mensch ist fremd – fast überall auf der Welt. Im Folgenden werden die Inhalte der einzelnen wissensbasierten Interventionseinheiten kurz beschrieben.

1. Wie leben die Menschen dieser Welt?

Hier sollten die Kinder zunächst eigene Erfahrungen zu der Frage zusammentragen, was für unterschiedliche Menschen sie bereits selbst kennen gelernt haben, und wie sich diese unterscheiden. Dann wurden anhand von DINA3-Fotografien acht Familien mit ihrem gesamten Hausrat aus unterschiedlichen Ländern¹³ vorgestellt (Bildmaterial aus: „So leben sie!“; Karpinski, 2001). Auch hier wurde zusammengetragen, welche kulturellen Besonderheiten und welche Vielfalt an Lebensweisen es gibt.

¹³ Die Bilder stellten Familien aus folgenden Ländern dar: Äthiopien, China, Deutschland, Indien, Kuwait, Mali, Südafrika, USA.

2. Die Welt-Schatzsuche

Hier wurden die Kinder in Gruppen eingeteilt und „bereisten“ mit ihrer Gruppe jeweils ein fremdes Land¹⁴. Die spielerische Expedition gab den Kindern anhand von Informationstafeln Einblicke in die geographischen und klimatischen Bedingungen, in die Flora und Fauna des jeweiligen Landes, ebenso wie in kulturelle Eigenheiten, die sich in Anpassung an die Bedingungen ihrer Umgebung entwickelt haben. Die Kinder lernten in Gruppenarbeit mit den PARTS-Materialien kulturelle Unterschiede kennen und ebenso den Ursprung und die Sinnhaftigkeit dieser Unterschiede. Zum Abschluss stellte jede Gruppe die Ergebnisse ihrer Expedition in der Klasse vor. Da die Welt-Schatzsuche die Verarbeitung vieler Informationen und die Erarbeitung komplexer Zusammenhänge erfordert, benötigte das Spiel mehr als eine PARTS-Einheit und wurde in zwei aufeinander folgenden Sitzungen durchgeführt.

3. Gemeinsamkeiten zwischen Kulturen

Hier lernten die Kinder anhand eines Beispiels von zwei Kindern aus verschiedenen Ländern, dass es neben den kulturellen Unterschieden auch viele Gemeinsamkeiten zwischen Menschen unterschiedlicher Kulturen gibt. Dazu wurde eine Kurzgeschichte vorgelesen, in der sich zwei fremde Jungen im Zug treffen und beginnen sich zu unterhalten. Dabei stellen sie fest, dass sie sich ziemlich ähnlich sind, obwohl der eine aus Kenia und der andere aus Deutschland kommt. Die PARTS-Teilnehmer mussten dann in einem Suchbild alle Gemeinsamkeiten herausfinden.

4. Kontakt zwischen Kulturen

Hier wurde zunächst zusammengetragen wie und warum es immer wieder dazu kommt, dass sich Menschen aus unterschiedlichen Ländern und Kulturen treffen. Im nächsten Schritt erfuhren die Kinder, dass es beim ersten Kontakt leicht zu Missverständnissen zwischen Menschen unterschiedlicher Kulturen kommen kann. Dazu wurde ein Begrüßungsspiel mit den Kindern gespielt, bei der jedes Kind eine kulturspezifische Art der Begrüßung auf einem Zettel zugelost bekam. In zufälligen Paaren begrüßten sich dann die Kinder mit ihrer Art der Begrüßung und stellten dabei unterschiedliche Passungen und Schwierigkeiten fest. Diese wurden gemeinsam ausgewertet. Im dritten Schritt wurde gemeinsam überlegt, was man tun kann, um den ersten Kontakt mit Menschen aus fremden Ländern zu erleichtern. Dazu wurde ein Beispielszenario über ein neues Kind in der Klasse aus Nigeria

¹⁴ Die Kinder „reisten“ während der Welt-Schatzsuche in folgende Länder: Algerien, Brasilien, Grönland, Nepal.

herangezogen. Die Kinder sammelten Ideen, was man tun kann, um die ersten Schwierigkeiten zu überbrücken, aufeinander zuzugehen und einen positiven Kontakt aufzubauen.

5. Jeder Mensch ist fremd – fast überall auf der Welt

In der letzten wissensbasierten Einheit wurde auf die Wahrnehmung von fremden Menschen eingegangen und die Begriffe Vorurteil, Stereotyp und Diskriminierung eingeführt. Dazu wurde mit den Kindern das Spiel „Wie betrachten wir Personen“ gespielt. Hier sollten die Kinder zunächst aus einer Auswahl aus drei unterschiedlich großen und unterschiedlich verpackten Paketen das auswählen, was sie zuerst auspacken würden. Es wurde eine Rangreihe der Pakete erstellt, und diese wurden der Reihe nach ausgepackt. Die Kinder stellten fest, dass die großen, schön verpackten Pakete nicht unbedingt etwas Schönes enthalten, und dass umgekehrt in den kleinen, weniger schön verpackten Paketen durchaus etwas Tolles stecken kann. Dann wurden die Kinder danach gefragt, aus welchen Gründen sie sich für bestimmte Pakete entschieden hatten. Hier lernten die Kinder, dass der erste Eindruck trügen kann, auch wenn dieser häufig zur Beurteilung (von Objekten wie von Personen) herangezogen wird. Diese Erkenntnis wurde auf ein Beispiel zur Auswahl von möglichen Spielkameraden übertragen. Hier wurde geklärt, wie man Personen betrachtet, welche Merkmale man beachtet und welche Schlüsse man daraus zieht. Außerdem wurde darauf eingegangen, dass diese ersten Eindrücke häufig von Vorurteilen geleitet sind und somit täuschen können.

Zum Abschluss wurden die Begriffe Vorurteile, Rassismus und Diskriminierung anhand von Textpassagen aus dem Buch „Papa, was ist ein Fremder“ (Jelloun, 2000) erläutert, und die Kinder sollten eigene Beispiele für die Begriffe zusammentragen.

Insgesamt wurden, über fünf Einheiten verteilt, fünf wissensbasierte Lektionen mit einer Gesamtinterventionsdauer von 180 Minuten durchgeführt. Einen Überblick zu den Einheiten mit wissensbasierten Interventionen bei PARTS liefert Tabelle 05.

Tabelle 05

Wissensbasierte Interventionen bei PARTS

Wissenseinheit	Inhalt	Ziel- fremdgruppen	Lektion	Dauer (Min.)
Wie leben die Menschen dieser Welt?	Vielfalt von Menschen und Lebensweisen der Welt aufzeigen	Äthiopien, China, Deutschland, Indien, Kuwait, Mali, Südafrika, USA.	2	35
Die Welt-Schatzsuche	Kulturelle Unterschiede basieren auf Unterschieden in den Umgebungsfaktoren.	Algerien, Brasilien, Grönland, Nepal.	4 und 5	35 und 15
Gemeinsamkeiten zwischen Kulturen	Es gibt viele Gemeinsamkeiten zwischen Menschen aus verschiedenen Ländern.	Kenia	5	25
Kontakt zwischen Kulturen	Ursachen für Kontakt zwischen Kulturen, mögliche Missverständnisse und Möglichkeiten positiven Kontakt herzustellen	Nigeria	10	35
Jeder Mensch ist fremd – fast überall auf der Welt	Der erste Eindruck bei der Beurteilung von Personen basiert meist nur auf dem Äußeren und kann leicht trügen. Personen erst besser kennen lernen und dann urteilen		14	35

Anmerkungen.

Förderung sozial-kognitiver Kompetenzen

Die dritte Säule von PARTS befasst sich mit der Förderung von zentralen sozial-kognitiven Fähigkeiten, die bei der Entwicklung und Veränderung von gruppenbezogenen sozialen Einstellungen eine Rolle spielen (s. Kapitel II.3. Empirische Studien). Hier wurden, auf Grund der zeitlichen Begrenzung von PARTS auf 15 Einheiten, auf Basis der Forschungsergebnisse drei zu fördernde sozial-kognitive Faktoren ausgewählt: 1. die Fähigkeit zur multiplen Klassifikation (Bigler & Liben, 1992, 1993), 2. die Fähigkeit zur Perspektivenübernahme und Empathie (Piaget & Weil, 1951; Aboud, 1988; Doyle & Aboud, 1995; Black-Gutman & Hickson, 1996) und 3. die Fähigkeit zur sozialen Problemlösung (Dovidio et al., 2004; Johnson & Johnson, 2000; Jugert, Kabak & Notz, 2006). Unterschiedliche Übungen und Spiele zur Förderung dieser drei Fähigkeiten wurden in das PARTS-Manual integriert. Auch hier wurde teilweise auf bereits bestehende Materialien zurückgegriffen, teilweise wurden aber auch eigene Übungen und Spiele entwickelt, um die einzelnen Bereiche zu fördern. Die Konzepte sowie Beispielübungen zur Förderung dieser drei Bereiche werden im Folgenden kurz vorgestellt.

1. Fähigkeit zur multiplen Klassifikation

Entsprechend der Idee, dass man Personen gleichzeitig auf verschiedenen Dimensionen einordnen können muss, um bei deren Beurteilung nicht auf Stereotype angewiesen zu sein (Bigler & Liben, 1992), wurde im Rahmen von PARTS mit unterschiedlichen Übungen versucht, die Fähigkeit zur multiplen Klassifikation (besonders mit sozialen Stimuli) zu verbessern. Angefangen bei sich selbst (z.B. Ich bin...-Arbeitsblatt), über regelmäßige Vergleiche mit Klassenkameraden auf unterschiedlichen Merkmalen beim Eröffnungsritual (Alle in einer Reihe) hin zur Klassifikation von anderen (Du bist...-Reportage, Gemeinsamkeiten finden) sollte die Fähigkeit zur multiplen Klassifikation spielerisch und mit unterschiedlichen Methoden gefördert werden. Dabei mussten die Kinder sich und andere immer wieder in unterschiedliche Kategorien einordnen. Damit sollte eine multidimensionale Betrachtung von Personen gefördert werden.

2. Fähigkeit zur Perspektivenübernahme und Empathie

Hier sollten die Kinder zunächst lernen, dass jeder Mensch eine eigene Sichtweise auf die Dinge hat, und diese sich von den Sichtweisen Anderer unterscheiden kann. Außerdem sollten die Kinder hier lernen, dass diese individuellen Sichtweisen gleichberechtigt nebeneinander stehen können und sollen

(vgl. Aboud, 1988; Piaget & Weil, 1951). Dies wurde den Kindern anhand von Kippbildern, die von verschiedenen Kindern unterschiedlich interpretiert werden, deutlich gemacht. Weiterhin sollten die Kinder üben, sich in die Situation von anderen hineinzuversetzen. Hierzu wurden Denkblasen-Arbeitsblätter bearbeitet, in denen die Kinder eintragen sollten, was Personen in einer dargestellten Interaktion denken oder sagen könnten. Außerdem wurden den Kindern Märchen aus einer völlig fremden Perspektive erzählt, und schließlich sollten sie die Akteure eines Rollenspiels beobachten und vermuten, welche Gedanken und Gefühle diese erleben.

3. Fähigkeit zum sozialen Problemlösen

Angelehnt an das Programm IKPL (Ich kann Probleme lösen! Beelmann, 2004) für Vorschulen und Kindergärten wurde ein Schema zur Lösung sozialer Probleme für Grundschüler entwickelt. Dieses Schema besteht aus drei Schritten, die jeweils mit einem Tier verknüpft und den Kindern mittels einer Fabel nahe gebracht werden. 1. Adler : Was ist das Problem? 2. Fuchs: Was gibt es für Lösungen? 3. Hirsch: Was ist die beste Lösung? Diese drei Schritte wurden mit den Kindern gemeinsam anhand der Fabel von Adler, Fuchs und Hirsch erarbeitet und dann auf verschiedene Problemszenarien, die alltägliche Konflikte der Kinder mit intergruppalen Kontexten thematisierten, angewendet.

Die Übungen und Spiele aus diesen drei Bereichen wurden gezielt auf die einzelnen PARTS-Lektionen aufgeteilt. Somit entstand ein abwechslungsreiches Programm, in dem aber immer wieder die gleichen Themen aufgegriffen wurden, um die nötige Intensität und die nötigen Wiederholungen für eine wirksame Förderung der einzelnen Fähigkeiten zu erreichen. Insgesamt wurden über acht Einheiten verteilt 13 Übungen zur Förderung von sozial-kognitiven Fähigkeiten mit einer Gesamtdauer von 200 Minuten durchgeführt. Davon wurden für die Fähigkeit zur multiplen Klassifikation 70 Minuten, für die Förderung von Perspektivenübernahmefähigkeiten 60 Minuten und für die Übungen zum sozialen Problemlösen 70 Minuten eingesetzt. Zusätzlich wurde in 14 Einheiten das Eröffnungsritual „Alle in einer Reihe“ zur Förderung der multiplen Klassifikationsfähigkeit durchgeführt, was insgesamt 70 Minuten Interventionszeit in Anspruch nahm. In der folgenden Tabelle 06 findet sich eine Übersicht der einzelnen Übungen und Spiele zur Förderung sozial-kognitiver Kompetenzen bei PARTS.

Tabelle 06

Förderung sozial-kognitiver Kompetenzen bei PARTS

Übung	sozial-kognitiver Faktor	Inhalt	Lektion	Dauer (Min.)
Alle in einer Reihe	multiple Klassifikationsfähigkeit	Kinder müssen sich und andere nach unterschiedlichen Merkmalen in immer neue Rangreihen bringen.	2 bis 15	5
Ich bin...-Arbeitsblatt	multiple Klassifikationsfähigkeit	Kinder ordnen sich allen Gruppen zu, denen sie angehören.	3	10
Du bist...-Reportage	multiple Klassifikationsfähigkeit	Kinder interviewen einen Freund oder Verwandten und erfragen, welchen Gruppen die Person angehört.	3 (Hausaufgabe)	10
Fabel von Adler, Fuchs und Hirsch	soziales Problemlösen	Kinder erlernen anhand einer Fabel ein Schema zum Lösen sozialer Probleme.	6	35
Problemlöseszenario „Fußball“ oder „Malkreide“	soziales Problemlösen	Kinder wenden das erlernte Schema zur Lösung von Problemen auf ein Szenario an.	7	20
Kippbilder	Perspektivenübernahme	Kinder sehen sich Kippbilder an und lernen, dass es unterschiedliche Sichtweisen auf dieselbe Sache geben kann.	7	10
Denkblasenbilder	Perspektivenübernahme	Kinder sollen sich überlegen, was Personen in einem Bild gerade denken könnten.	7	10

Märchen aus der Sicht des Bösewichts	Perspektiven-übernahme	Kinder lesen Märchen aus der Sicht des Bösewichts und lernen, dass jede Situation zwei Seiten hat.	8	20
Gemeinsamkeiten finden	multiple Klassifikationsfähigkeit	Kinder versuchen in Kleingruppen möglichst viele Gemeinsamkeiten zu finden.	9	20
Rollenspiel	Perspektiven-übernahme und Empathie	Kinder beobachten eine kurze vorgespielte Szene und sollen angeben, was die beiden Personen gedacht und gefühlt haben.	11	20
Ich mag alle die...	multiple Klassifikationsfähigkeit	Im Stuhlkreis wird ein Merkmal genannt, worauf alle Merkmalsträger ihren Sitzplatz wechseln müssen. Wer übrig bleibt, gibt das nächste Merkmal vor.	11	15
Arbeitsblätter MK	multiple Klassifikationsfähigkeit	Kinder müssen Lücken in Matrizen füllen und dazu die relevanten Merkmalsdimensionen erkennen.	11 und	5 +
Gleiche Merkmale finden	multiple Klassifikationsfähigkeit	Kinder bewegen sich durch den Raum und müssen sich nach unterschiedlichen Merkmalen zu gleichen Gruppen zusammenfinden.	12	10
Problemlöseszenario „Gameboy“	soziales Problemlösen	Kinder wenden das erlernte Schema zur Lösung von Problemen auf Szenario an.	13	15

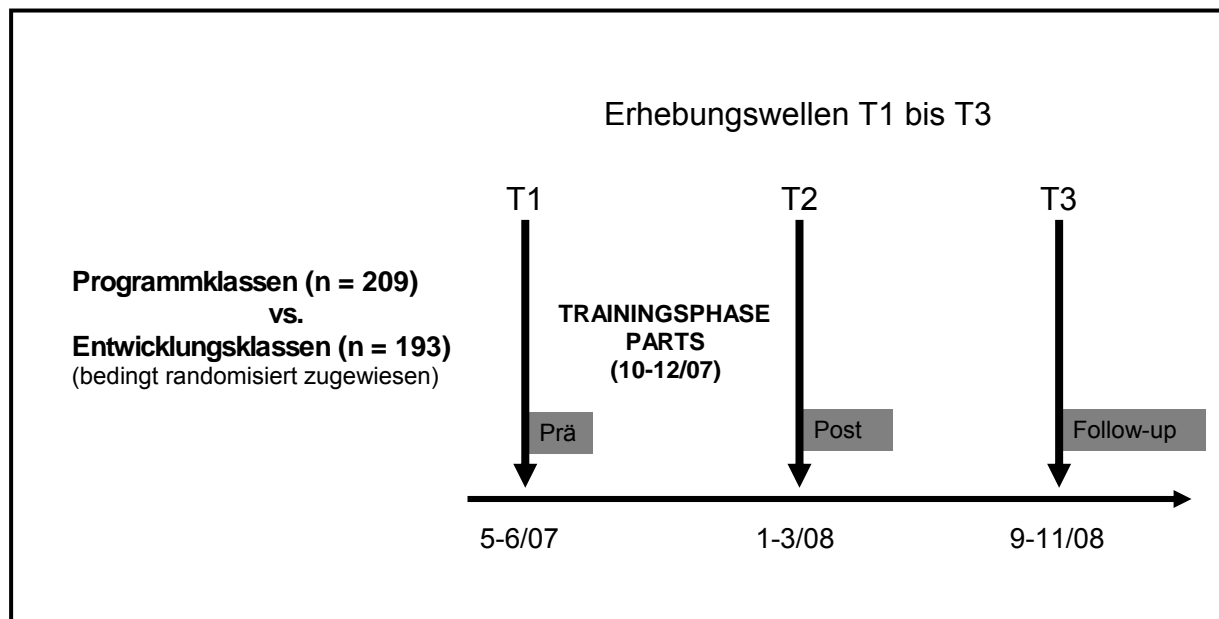
Anmerkungen.

III.3. *Design der Untersuchung*

Zur Evaluation von PARTS wurde ein Interventions-Kontrollgruppen-Design mit drei Messzeitpunkten (Prätest, Posttest und Follow-up) gewählt. Dabei wurden die Kinder auf Klassenebene bedingt randomisiert zu den beiden Untersuchungsgruppen zugeordnet. Hierzu wurden zunächst Klassenpaarlinge nach den Kriterien Stadtschule versus Landschule sowie Klassengröße gebildet. Von diesen Paarlingen wurde jeweils eine Klasse der Interventionsgruppe und eine Klasse der Kontrollgruppe zugeordnet. Die Zuordnung erfolgte zufällig per Auslosung. Die bedingte Randomisierung auf Klassenebene erfolgte, da eine rein zufällige Zuordnung der einzelnen Kinder zu den Untersuchungsgruppen die Integration des Interventionsprogramms in den Schulunterricht unverhältnismäßig erschwert hätte, was die Kooperationsbereitschaft der beteiligten Schulen deutlich gemindert hätte. Da alle Klassen unter den gleichen bekannten Bedingungen randomisiert zugeordnet wurden, und da die Teilnehmerzahl ausreichend groß ist, kann trotzdem von statistisch unabhängigen Untersuchungsgruppen ausgegangen werden (Cook & Campbell, 1979).

Ein derartiges Untersuchungsdesign mit einer verhältnismäßig umfangreichen Stichprobe ermöglicht es, zuverlässige Aussagen über die Wirksamkeit der Intervention zu machen. Gleichzeitig werden dabei diverse Validitätsgefährdungen für die Aussagekraft der Studie, wie zufällige Gruppenunterschiede, unvorhergesehene Einflüsse, Reifungseffekte, Testeffekte, Selektionseffekte oder eine geringe Repräsentativität der Stichprobe, vermieden (Cook & Campbell, 1979). Schließlich ermöglicht die Messung über 3 Zeitpunkte die längerfristige Wirkung von PARTS einschätzen zu können, was die zeitliche Stabilität der Interventionseffekte absichert. Abbildung 03 zeigt das Design der PARTS-Studie.

Abbildung 03

Design der PARTS-Studie

III.4. Die Stichprobe

Um zuverlässige Aussagen über die Wirksamkeit von PARTS auf deutsche Kinder der dritten Klassen machen zu können, ist eine ausreichend große und repräsentative Stichprobe erforderlich. Diese Stichprobe wurde mit erheblichem Aufwand rekrutiert und gepflegt, um die Drop-out-Rate so gering wie möglich zu halten. In den nächsten Abschnitten wird ausführlich auf die Rekrutierung der Stichprobe, deren Entwicklung über die drei Messzeitpunkte hinweg, die Stichprobenpflege sowie die notwendigen Fallausschlüsse für die Analysen zur Evaluation von PARTS eingegangen. Schließlich wird die Stichprobe anhand einiger sozio-demographischer Variablen beschrieben und die Datenlage für die Evaluation von PARTS zusammengefasst.

III.4.1. Rekrutierung der Stichprobe

Um das DFG-Projekt „Förderung von Intergruppenbeziehungen bei Kindern“ mit einer ausreichend großen, repräsentativen und gut erreichbaren Stichprobe zu versorgen, stellten die Projektleiter Prof. Dr. Beelmann und Prof. Dr. Noack¹⁵ die Projektidee im Frühjahr 2007 auf zwei Schulleiterkonferenzen des Schullandkreises Gotha/Bad Langensalza vor. Zuvor wurde das Projekt beim Thüringer Kultusministerium präsentiert und genehmigt.

Bei der ersten Schulleiterkonferenz handelte es sich um den „Landkreis Gotha/ Stadt Gotha und Waltershausen“. In diesem Landkreis gab es zu diesem Zeitpunkt 27 Grundschulen. Im Rahmen der Schulleiterkonferenz konnten sich die Schulleiter im Anschluss an eine etwa einstündige Projektpräsentation in Interessentenlisten eintragen, wenn sie sich für das Projekt interessierten und sich prinzipiell vorstellen konnten, dass dieses an ihren Schulen verwirklicht werden sollte. Hier haben sich 13 Schulleiter eingetragen und ihr Interesse bekundet.

¹⁵ Das Projekt Förderung von Intergruppenbeziehungen bei Kindern war ein Kooperationsprojekt zweier Arbeitsgruppen. Die Arbeitsgruppe um Prof. Dr. Noack lieferte dabei vor allem Unterstützung bei der Entwicklung der Messinstrumente, der Befragung der Kinder in der Kontrollgruppe sowie der Aufbereitung der entsprechenden Daten.

Bei der zweiten Rektorenkonferenz, die besucht wurde, handelt es sich um den Landkreis „Unstrut-Hainich“. In diesem Landkreis gab es zu diesem Zeitpunkt 24 staatliche Grundschulen und vier Grundschulen in freier Trägerschaft, insgesamt also 28 Grundschulen. Von diesen 28 Schulen haben sich 17 SchulleiterInnen in die Interessentenliste eingetragen.

Anschließend wurden in den nächsten Wochen alle interessierten Schulen nochmals telefonisch kontaktiert. Dies geschah nach vorheriger Absprache mit den Schulleitern, um diesen die Möglichkeit zu geben, das Projektvorhaben mit ihren Lehrern abzusprechen. Von den insgesamt 30 zunächst interessierten Schulen (also etwa 55% aller Grundschulen aus dem gesamten Schullandkreis) sagten sieben Schulen während des Telefonats ab. Vier dieser Schulen sagten aus Zeitgründen ab, an zwei Schulen hatten die Klassenlehrer kein Interesse an dem Projekt, und eine Schule sagte ohne Angabe von Gründen ab. Somit befanden sich zu diesem Zeitpunkt noch 23 Schulen im Pool für die Projektstichprobe. Aus Kapazitätsgründen waren wir leider nicht in der Lage, allen 23 Schulen die Teilnahme am Projekt zu ermöglichen. Daher wurden 8 Schulen per Losverfahren aus dem Pool der Schulen ausgeschlossen. Somit verblieben 15 Schulen mit insgesamt 24 teilnehmenden Klassen in der Stichprobe, was einer Quote von 27% der Schulen des Schullandkreises entspricht. Zehn dieser Schulen lagen im Landkreis Gotha und fünf Schulen im Landkreis Unstrut-Hainich.

Im Anschluss an die Auswahl der Schulen wurden nun die teilnehmenden Klassen den beiden Untersuchungsgruppen *Interventionsgruppe* und *Kontrollgruppe* zugewiesen. Dazu wurden zunächst Klassenpaarlinge nach den Kriterien Stadt vs. Land und der Größe der Schule gebildet. Aus diesen Paarlingen wurden dann jeweils eine Klasse der Interventionsgruppe und eine Klasse der Kontrollgruppe zufällig zugeordnet (s.o.). Somit ergab sich die folgende Aufteilung der Schulen und Klassen zu beiden Gruppen: Vier Schulen nahmen mit jeweils einer Klasse ausschließlich an der Interventionsgruppe teil. Drei Schulen beteiligten sich mit jeweils einer Klasse nur an der Kontrollgruppe und acht Schulen nahmen mit insgesamt 17 Klassen sowohl an der Interventions- als auch an der Kontrollgruppe teil. Insgesamt befanden sich damit jeweils 12 Klassen mit einer Gesamtanzahl von $N = 502$ Kindern in beiden Gruppen im Pool der Stichprobe. Davon sollten 247 Kinder der Interventionsgruppe und 255 Kinder der Kontrollgruppe angehören. Diese Anzahl berücksichtigt allerdings noch nicht die Teilnahmegenehmigung der Kinder durch deren Eltern. Die

Teilnahmegenehmigungen wurden an die Schulen versandt und an die Kinder der teilnehmenden zweiten Klassen ausgeteilt. Alle Eltern hatten die Möglichkeit, sich bei Rückfragen an uns zu wenden, was auch einige Eltern in Anspruch genommen haben. Insgesamt wurde 430 Kindern die Genehmigung zur Teilnahme an der Studie erteilt. Das entspricht einem Anteil von etwa 86% der Kinder aus allen teilnehmenden Klassen. Aus welchen Gründen 14% der Eltern ihre Kinder nicht an der Untersuchung teilnehmen ließen, bleibt leider Spekulationen überlassen. Nur sehr wenige Eltern gaben dazu eine Rückmeldung und begründeten ihre Entscheidung dann meist mit der Sorge um zu viel Unterrichtsausfall durch die Teilnahme der Kinder am Projekt. Die übrigen $N = 430$ Kinder teilten sich mit $n = 217$ Kindern zur Interventionsgruppe und $n = 213$ Kindern in der Kontrollgruppe auf¹⁶.

III.4.2. Entwicklung der Stichprobe

Zur zweiten Befragungswelle konnte ein Großteil der teilnehmenden Kinder wieder befragt werden. Lediglich fünf Kinder sind aus der Stichprobe herausgefallen, da diese mit ihren Familien in weit entfernte Regionen (meist andere Bundesländer) umgezogen und somit für uns nicht mehr erreichbar waren. Weitere vier Kinder konnten auf Grund von Umzügen und Schulwechseln nicht an der zweiten Messung teilnehmen, zur dritten Messung allerdings wieder erreicht werden. Da bei diesen Kindern jedoch der Posttest fehlte, wurden sie ebenso wie die fünf Drop-out-Kinder von den Analysen zur Evaluation von PARTS ausgeschlossen.

Das bedeutet, die Studie hat vom ersten zum zweiten Messzeitpunkt eine Retention Rate von 97.9%. Keine Familie hat sich dazu entschlossen, die Teilnahme am Projekt abubrechen, was einen großen Erfolg für unser Projekt darstellt.

Die Retention Rate betrug zur dritten Erhebungswelle ebenfalls 97.9%. Lediglich vier Kinder aus der Grundgesamtheit vom ersten Messzeitpunkt nahmen an der dritten Befragungswelle nicht teil. Drei davon konnten auf Grund von Umzügen nicht wieder erreicht werden. Ein Kind verweigerte zur dritten Messung die

¹⁶ Allerdings wurden in einer Schule die Klassen nach der 2. Klasse, also nach der Zuordnung zu Interventionsgruppe und Kontrollgruppe, auf Grund geringer Schülerzahlen zusammengelegt. Die neue Gesamtklasse wurde in der Interventionsgruppe untergebracht. Daher fällt die Schülerzahl in der Interventionsgruppe letztendlich etwas größer aus, als in der Kontrollgruppe (s.u., Kapitel III.4.5. Beschreibung der Evaluationsstichprobe).

Teilnahme an der Befragung. Alle vier Kinder wurden für die weiteren Analysen ausgeschlossen. Eine Übersicht über die Entwicklung der Stichprobe findet sich in Tabelle 07.

Tabelle 07

Teilnahme und Ausfallraten an den PARTS-Untersuchungen, bezogen auf die Teilnehmer zu t1

Messzeitpunkt	N gesamt/ N untersucht	Drop-outs/ Missings	Gesamtausfallrate in Prozent
T1	430/430	0	0
T2	425/421	5/4	2.1 %
T3	422/421	3/1	2.1 %

Anmerkungen. Die Ausfallrate in % bezieht sich auf die fehlenden Daten pro Messzeitpunkt insgesamt (Drop-outs + Missings).

III.4.3. Stichprobenpflege

Allgemein verlief die Zusammenarbeit mit den Schulen und den Kindern sehr gut. Terminliche Absprachen zu den Datenerhebungen ließen sich gut realisieren, und auch die Integration des Programms in den normalen Schulunterricht verlief hervorragend. Sowohl Lehrer – alle teilnehmenden Klassenlehrer – als auch Eltern – über 90% der teilnehmenden Familien – beteiligten sich an unseren Befragungen. Die Kinder hatten an den Befragungen und am Programm viel Spaß und freuten sich immer, wenn unsere Mitarbeiter in ihre Schulen kamen. Dies alles spricht für eine sehr gute Zusammenarbeit zwischen Eltern, Schulen und Projektmitarbeitern und für das große Interesse an dem Projekt, sowie für die Freude der Kinder, daran teilnehmen zu können.

Im Rahmen der Stichprobenpflege und als Anreiz für die Teilnahme an der Studie wurde sowohl den teilnehmenden Familien, als auch den Schulen, für die Teilnahme an allen drei Untersuchungswellen eine Aufwandsentschädigung von 50,- € gezahlt. Hierzu wurden die Bankverbindungen der Familien und der Klassenkassen der beteiligten Klassen erfragt. Die Aufwandsentschädigung wurde nach der Teilnahme der Familien an der zweiten Untersuchungswelle ausgezahlt. Eine weitere

Maßnahme der Stichprobenpflege war ein Weihnachtsbrief, den jede beteiligte Familie und die Schulen vom Projektteam vor dem Weihnachtsfest 2007 erhielten¹⁷.

III.4.4. Fallausschlüsse für die Evaluationsstichprobe

Da vier Kinder zwar die Interventionsklassen besuchten, aber aus Krankheitsgründen nicht am Programm teilnehmen konnten, wurden diese Kinder aus den Auswertungen ausgeschlossen. Weiterhin wurden die Daten von zwei Kindern aus den Analysen entfernt, welche fünf oder mehr Einheiten des Interventionsprogramms auf Grund von Erkrankungen versäumten, da diesen Kindern nicht alle Inhalte des Programms vermittelt werden konnten und zudem die notwendige Programmintensität nicht gewährleistet war.

Da PARTS als Präventionsprogramm für die deutsche Majorität konzipiert wurde, wurden außerdem neun Kinder mit Migrationshintergrund (wenn das Kind selbst oder ein Elternteil nicht in Deutschland geboren wurde) aus der Stichprobe für die Evaluation ausgeschlossen, um mögliche systematische Verzerrungen zu vermeiden. Somit ergibt sich schließlich zur Evaluation des Interventionsprogramms eine Stichprobe von $N = 402$ Kindern. Diese 402 Kinder gehören zu 52 % (209 Kinder) der Interventionsgruppe und zu 48 % (193 Kinder) der Kontrollgruppe an.

Dieser Unterschied in der Gruppengröße entstand, wie oben bereits erwähnt, nach dem Prätest durch die Zusammenlegung von zwei kleineren Klassen. Die neue Gesamtklasse wurde der Interventionsgruppe zugeordnet. Eine Übersicht zu sämtlichen Fallausschlüssen und Ausfällen aus der Evaluationsstichprobe der PARTS-Studie findet sich in Tabelle 08.

¹⁷ Die Weihnachtsbriefe finden sich im Anhang C.

Tabelle 08

Fallausschlüsse und Ausfälle zur Evaluationsstichprobe

Kriterium	Anzahl	Gruppe		Geschlecht		Ursachen
		IG	KG	J	M	
Drop-outs nach t1	5	3	2	1	4	5 U
Drop-outs nach t2	3	2	1	1	2	3 U
Missings zu t2	4	2	2	2	2	3 U & 1 S
Missings zu t3	1	1		1		1 V
Interventionsgruppe, aber nicht an PARTS teilgenommen	4	4		2	2	4 K
Interventionsgruppe, aber 5 oder mehr Einheiten verpasst	2	2		1	1	2 K
Migrationshintergrund	9	4	5	2	7	3 GK & 6 GE

Anmerkungen. IG = Interventionsgruppe, KG = Kontrollgruppe, J = Junge, M = Mädchen; U = Umzug, S = Schulwechsel, V = Verweigert, K = Krank, GK = Geburtsland des Kindes, GE = Geburtsland der Eltern; Die Kriterien zum Fallausschluss wurden in der Reihenfolge der Tabelle von oben nach unten angewendet.

III.4.5. Beschreibung der Evaluationsstichprobe**Soziodemographische Variablen der Untersuchungsteilnehmer**

Die $N = 402$ Teilnehmer, deren Daten für die weitere Auswertung berücksichtigt werden, teilen sich zu $n = 209$ (52 %) Kindern in die Interventionsgruppe und $n = 193$ (48 %) Kindern in die Kontrollgruppe auf. Dabei sind die Geschlechter nahezu gleich verteilt: $n = 200$ Jungen und $n = 202$ Mädchen nahmen an der Untersuchung teil. Eine ähnliche Geschlechterverteilung findet sich auch innerhalb der beiden Untersuchungsgruppen (s.u. Tabelle 09). Die Kinder waren zum ersten Messzeitpunkt im Durchschnitt etwa acht Jahre und fünf Monate alt und besuchten die zweite Grundschulklasse. Das Trainingsprogramm begann vier Monate nach dem Prätest im Herbst 2007. Die teilnehmenden Kinder waren zum

Trainingsbeginn also in der dritten Klasse und im Durchschnitt acht Jahre und neun Monate alt. Etwa zwei Monate nach Beendigung des Trainingsprogramms fanden die ersten Erhebungen für den Posttest im zweiten Schulhalbjahr des Schuljahres 2007/2008 statt. Das Follow-up folgte schließlich acht Monate später im Herbst 2008. Nun waren die Kinder im Schnitt neun Jahre und neun Monate alt und besuchten die vierte Klasse. Die Alters- und Geschlechterverteilung in den Untersuchungsgruppen ist vergleichbar (Alter zu t1: $F [1, 401] = 0.19, p = 0.664$; Geschlecht: $F [1, 401] = 0.1, p = 0.845$) und findet sich zur besseren Übersicht in Tabelle 09.

Auch auf anderen, möglicherweise relevanten, Variablen unterschieden sich die Teilnehmer der beiden Untersuchungsgruppen nicht. So ist die Verteilung des Entwicklungskontextes der Kinder, also der Anteil der Kinder, die in ländlicher bzw. in städtischer Umgebung aufwachsen, in beiden Untersuchungsgruppen gleich ($F [1, 400] = 0.04, p = 0.838$). Der Anteil der „Städter“ ist allerdings insgesamt deutlich geringer und liegt bei etwa einem Viertel der Untersuchungsteilnehmer ($n = 94, 23.4 \%$). Weiterhin finden sich keine bedeutsamen Unterschiede zwischen der Interventionsgruppe und der Kontrollgruppe in Bezug auf die mit dem Raven-Matrizen test gemessene Intelligenz der Kinder ($F [1, 397] = 1.41, p = 0.237$). Hier liegen die Kinder insgesamt bei einem Mittel von $M = 3.32$ bei einer Skala von 0 bis 8.

Tabelle 09

Geschlecht und Alter der Evaluationsstichprobe über die Zeit

Gruppe	Geschlecht		Alter T1	Alter PARTS	Alter T2	Alter T3
	J	M	M (SD)			
Interventionsgruppe n = 209	103	106	101 (4.75)	105 (4.7)	109 (4.78)	117 (4.8)
Kontrollgruppe n = 193	97	96	101 (4.25)	105 (4.19)	109 (4.28)	116 (4.21)
gesamt N = 402	200	202	101 (4.51)	105 (4.46)	109 (4.54)	117 (4.53)

Anmerkungen. J = Junge, M = Mädchen; M (SD) = Mittelwert Alter in Monaten auf ganze Monate gerundet (Standardabweichung). Alter PARTS bezieht sich auf das Alter der Kinder zu Beginn des Trainingsprogramms.

Soziodemographische Variablen der Eltern der Untersuchungsteilnehmer

Von den an PARTS teilnehmenden Kindern nahmen zum großen Teil auch deren Eltern an den Befragungen teil. So konnten zum ersten Messzeitpunkt die Daten von 380 (94.8 %) Müttern und 337 (83.8 %) Vätern erfasst werden.

Bei den Müttern, deren Angaben vorlagen, handelte es sich zum überwiegenden Teil um die leiblichen Mütter (97.11 %). Ein geringer Anteil (1.31 %) waren sonstige erziehungsberechtigte Personen, also etwa Großmütter, oder neue Lebenspartnerinnen des Vaters. Sechs der befragten Personen (1.58 %) machten keine Angaben zu dieser Frage. Die Mütter waren zu t1 zwischen 25 und 74 Jahren und im Durchschnitt 36 Jahre alt ($SD = 5.00$). Gut zwei Drittel der Mütter ist (war) verheiratet (72.11 %) und 16.84 % lebten in einer eheähnlichen Lebensgemeinschaft. 9.47 % der Mütter waren allein erziehend und 1.58 % machten keine Angaben zu ihren Lebensverhältnissen.

Lediglich drei Mütter (0.89 %) gaben bei der Frage nach dem höchsten Bildungsabschluss keinen an. 2.37 % hatten ihre Schullaufbahn nach der achten Klasse beendet und 5.26 % hatten den Hauptschulabschluss erreicht. Der Großteil der Mütter gab als höchsten Schulabschluss den Realschulabschluss an (69.74 %), während nur 7.63 % das Abitur oder Fachabitur erreichten. Schließlich gaben 13.68 % der Mütter an, einen Hochschul- oder Fachhochschulabschluss erworben zu

haben. Der Rest machte keine Angaben zu dieser Frage. Zur Frage nach ihrer derzeitigen Arbeit gaben 39.21 % der Mütter an, voll berufstätig zu sein, 33.68 % waren teilzeitbeschäftigt und 9.74 % waren arbeitslos. 6.05 % der Mütter gaben an Hausfrau zu sein und 4.74 % waren selbstständig tätig, während 5.53 % der Mütter sonstigen beruflichen Tätigkeiten nachgingen. 1.05 % der Mütter machte hierzu keine Angaben.

Bei den teilnehmenden Vätern handelte es sich um 89.02 % leibliche Väter und 7.12 % sonstige erziehungsberechtigte Personen. Der Rest der Väter (3.86 %) machte keine Angaben zur Beziehung zum teilnehmenden Kind. Die Väter waren zum ersten Messzeitpunkt zwischen 24 und 59 Jahren alt, wobei der Mittelwert bei 38.6 Jahren ($SD = 5.2$) lag. Zur Frage nach den Familienverhältnissen gaben 80.12 % der Väter an verheiratet zu sein, 16.02 % lebten in einer eheähnlichen Lebensgemeinschaft und 2.08 % der Väter waren allein erziehend. Der Rest (1.78 %) machte keine Angaben zu den Familienverhältnissen.

Nur zwei Väter (0.59 %) gaben an, keinen Schulabschluss zu haben, 5.34 % absolvierten erfolgreich die achte Klasse, während 5.05 % den Hauptschulabschluss erreichten. Der größte Teil der Väter (62.61 %) gab den Realschulabschluss als höchsten Bildungsabschluss an, 8.9 % absolvierten erfolgreich die (Fach-) Hochschulreife und 17.51 % gaben ein abgeschlossenes Hochschulstudium/Fachhochschulstudium als höchsten Bildungsabschluss an. Der größte Teil der Väter war voll berufstätig (79.23 %), während nur 1.78 % der Väter teilzeitbeschäftigt war. 9.79 % der Väter arbeiteten selbstständig, 5.34 % waren arbeitslos und 0.59 % gaben an Hausmänner zu sein. Die restlichen Väter (2.97 %) gingen sonstigen Tätigkeiten nach oder machten keine Angaben (0.3 %).

Insgesamt handelt es sich augenscheinlich bei der vorliegenden Stichprobe um einen repräsentativen Ausschnitt der Bevölkerung im eher ländlichen Thüringen. Auch wenn man die soziodemographischen Angaben der Eltern getrennt für die Interventions- und Kontrollgruppe betrachtet, zeigen sich in beiden Gruppen sehr ähnliche Verteilungen, wie in Tabelle 10 und Tabelle 11 ersichtlich wird. Somit kann von soziodemographisch äquivalenten Untersuchungsgruppen ausgegangen werden.

Tabelle 10

Vergleich soziodemographischer Variablen der Mütter nach Interventions- und Kontrollgruppe in Prozent

		Interventions- gruppe	Kontrollgruppe
Beziehung zum Kind	leibliche Mutter	94.7%	88.6%
	sonstige	0.5%	2.1%
Alter in Jahren zu t1			
Mittel (SD)		35.90 (4.54)	36.28 (5.48)
Lebensverhältnisse	Ehe	67.9%	68.4%
	eheähnliche		
	Lebensgem.	19.1%	12.4%
	alleinerziehend	8.1%	9.8%
höchster			
Bildungsabschluss	keiner	1.4%	0.0%
	achte Klasse	2.4%	2.1%
	Hauptschule	3.3%	6.7%
	Realschule	65.1%	66.8%
	Gymnasium	8.6%	5.7%
	Hochschule/		
	Fachhochschule	15.3%	10.4%
Arbeit	Vollzeit	40.7%	33.2%
	Teilzeit	28.7%	35.2%
	selbstständig	6.2%	2.6%
	Hausfrau	7.2%	4.1%
	arbeitslos	8.6%	9.8%

Anmerkungen. Angaben in Prozent. Die zu 100% fehlenden Anteile sind fehlende Werte in Bezug zur Grundgesamtheit.

Tabelle 11

Vergleich soziodemographischer Variablen der Väter nach Interventions- und Kontrollgruppe in Prozent

		Interventions- gruppe	Kontrollgruppe
Beziehung zum Kind	leiblicher Vater	77.5%	71.5%
	sonstige	7.2%	4.7%
Alter in Jahren zu t1			
Mittel (SD)		38.33 (5.00)	38.92 (5.43)
Lebensverhältnisse	Ehe	67.5%	66.8%
	eheähnliche		
	Lebensgem.	16.3%	10.4%
	alleinerziehend	1.9%	1.6%
höchster			
Bildungsabschluss	keiner	0.0%	1.0%
	achte Klasse	4.8%	4.1%
	Hauptschule	4.3%	4.1%
	Realschule	51.2%	53.9%
	Gymnasium	10.5%	4.1%
	Hochschule/		
	Fachhochschule	17.2%	11.9%
Arbeit	Vollzeit	67.5%	65.3%
	Teilzeit	1.9%	1.0%
	selbstständig	9.1%	7.3%
	Hausmann	1.0%	0.0%
	arbeitslos	5.7%	3.1%

Anmerkungen. Angaben in Prozent. Die zu 100% fehlenden Anteile sind fehlende Werte in Bezug zur Grundgesamtheit.

III.4.6. Beschreibung der Datenlage

In der untenstehenden Tabelle 12 wird aufgeschlüsselt, wie groß der Anteil der vollständig vorliegenden und für die Evaluation nutzbaren Daten der Untersuchungsteilnehmer ist. Bei der Angabe der Kinderdaten beschreiben die zu 100 % fehlenden Prozente den Anteil an Kinderdatensätzen, die unvollständig sind. Hier wurden einzelne Instrumente nicht erhoben oder einzelne Kinder konnten an der Einzel- oder der Gruppenbefragung nicht teilnehmen. Fehlende Daten bei den Elterndaten beruhen auf mangelndem Rücklauf aus den Familien. Fehlende Lehrerdaten beruhen auf Auslassungsfehlern der Lehrer bei der Datenerhebung. Die Beteiligung der Teilnehmer an den drei Untersuchungen kann mit einem nutzbaren Datenanteil von 80.6 % (Väterdaten zu t3) bis 100 % (Lehrerdaten zu t2) als sehr gut bezeichnet werden. Dies kann wohl auch auf die aufwändige Stichprobenpflege (s.o.) zurückgeführt werden. Die Beteiligung der Väter an den Elternbefragungen fällt dabei etwas geringer aus als die übrigen Werte. Mit einer Teilnahmequote von über 80 % ist aber auch hier die Datenlage noch sehr zufriedenstellend. Zum dritten Messzeitpunkt wurden keine Daten von den beteiligten Klassenlehrern erfragt. Dies geschah, um die Belastung der Lehrer durch die Untersuchung so gering wie möglich zu halten, und somit eine gute weitere Zusammenarbeit mit den Schulen und Lehrern zu gewährleisten.

Tabelle 12

Datenlage in der Evaluationsstichprobe

Daten		T1	T2	T3
Kinderdaten		398 (99 %)	401 (99.8 %)	402 (100 %)
Elterndaten	Mütter	380 (94.8 %)	386 (96.3 %)	366 (91 %)
	Väter	337 (83.8 %)	344 (85.6 %)	324 (80.6 %)
Lehrerdaten		397 (99 %)	402 (100%)	-

Anmerkungen. Lehrerdaten wurden zu t3 nicht erfragt.

III.5. Durchführung der Untersuchungen

Die Untersuchungen insgesamt sowie die Anwendung der einzelnen Messinstrumente verliefen hoch standardisiert, um so wenig Störvarianz wie möglich zu erzeugen. Im Folgenden wird der Ablauf der einzelnen Untersuchungswellen beschrieben.

III.5.1. Der Prätest T1

Die erste Befragung der an der Studie teilnehmenden Kinder fand im Sommer 2007 statt. Der Prätest begann am 08. und 09.05.2007 in der Grundschule Hörselgau und endete am 09. und 10.07.2007 in der Grundschule Dachwig. Somit ergibt sich ein Erhebungszeitraum von knapp neun Wochen.

Die Befragungen verliefen dabei nach einem einheitlichen Schema und bestanden aus zwei Teilen: einer Gruppenbefragung und den Einzelbefragungen.

Zunächst wurden in der Gruppenbefragung alle teilnehmenden Kinder der Schulklasse im Klassenverband mit einem für Kinder geeigneten Fragebogen befragt. Der Versuchsleiter stellte sich und sein Team als Mitarbeiter der Universität Jena vor und erklärte, sich dafür zu interessieren, was Kinder denken. Weiterhin wurden die Kinder über den weiteren Ablauf der Untersuchung kurz aufgeklärt. Diejenigen Kinder, welche keine Teilnahmegenehmigung ihrer Eltern erbringen konnten, wurden während der Befragung in eine andere Klasse gebracht oder erhielten von der Lehrkraft eine alternative Aufgabe zur Bearbeitung. Während der Befragung waren sowohl der Versuchsleiter mit dem Erhebungsteam als auch die zuständige Lehrkraft (meist die Klassenlehrerin) anwesend. Diese erhielt einen Umschlag mit Fragebögen zur Einschätzung des Verhaltens der teilnehmenden Kinder mit der Bitte, die Fragebögen auszufüllen und gemeinsam mit den eingesammelten Elternfragebögen mittels eines frankierten Rückumschlags an die Universität zu senden.

Dann bekam jedes verbliebene Kind einen geschlechtsspezifischen Fragebogen ausgeteilt. Der Fragebogen wurde vom Versuchsleiter erläutert. Die ersten einfachen demographischen Angaben wurden erfragt und die Kinder trugen ihre Antworten in die Fragebögen ein. Dann wurde das Antwortschema der nun folgenden Fragen mit Hilfe einer standardisierten Instruktion und anhand eines

Beispielitems deutlich gemacht. Die Kinder erhielten die Gelegenheit Fragen zu stellen, um Unklarheiten auszuräumen. Sobald alle Kinder das Antwortschema verstanden hatten, begann der Versuchsleiter mit dem Vorlesen der einzelnen Items. Dabei wurde darauf geachtet, jedes Item zweimal vorzulesen und langsam genug vorzugehen, um sicherzustellen, dass die Geschwindigkeit auch für langsamere Kinder angemessen war. Die Kinder wurden zuvor außerdem darauf hingewiesen, dass sie sich bei Verständnisschwierigkeiten jeder Zeit melden könnten. In solchen Fällen half ein Mitarbeiter des Erhebungsteams, diese Schwierigkeiten auszuräumen. Dieses Angebot wurde des Öfteren auch in Anspruch genommen, so dass davon ausgegangen werden kann, dass die Kinder, teilweise mit Unterstützung, alle Fragen verstanden haben.

Im Anschluss an den Fragebogen erhielten die Kinder noch zwei zusätzliche Tests, die in der Gruppe bearbeitet wurden. Zunächst wurde der Raven-Matrizen-Test präsentiert. Die Matrizen wurden den Kindern ausgeteilt und außerdem als Folie auf einem Overheadprojektor an eine freie Wand projiziert. Daraufhin wurde die ebenfalls standardisierte Instruktion verlesen und die Möglichkeit gegeben, Fragen zu klären. Im Anschluss konnten die Teilnehmer selbstständig die Aufgaben lösen. Auch hier stand das Erhebungsteam während der Bearbeitung zur Klärung von Fragen bereit und achtete darauf, dass die Kinder sich voll und ganz auf ihre Aufgaben konzentrierten und nicht beim Sitznachbarn abguckten. Nach Beendigung des Raven-Matrizen-Tests erhielten die Kinder noch das Intended Behaviour Measure. Auch dieses Maß wurde standardisiert instruiert und das Antwortschema anhand einer Folie erklärt. Nachdem die Teilnehmer auch dieses Maß mit der Unterstützung der Versuchsleiter bearbeitet hatten, konnten die ausgegebenen Tests und Fragebögen wieder eingesammelt werden. Dabei wurden alle eingesammelten Bögen mit einem individuellen Code versehen, der zuvor bereits auf Code-Listen festgelegt worden war. Dies war sowohl für die Zuordnung der individuellen Datensätze während der längsschnittlichen Erhebung erforderlich, als auch zur Einhaltung der Datenschutzbestimmungen notwendig. Nachdem alle Daten der Kinder eingesammelt und codiert worden waren, wurde die Gruppenbefragung abgeschlossen. Die komplette Gruppenbefragung dauerte dabei je nach Klasse zwischen 50 und 65 Minuten.

Im Anschluss an die Gruppenbefragung fanden die Einzelbefragungen statt. Die Einzelinterviews wurden jeweils von einem Versuchsleiter mit einem Kind

durchgeführt. Die Schulen stellten hierfür je nach Kapazität eine Anzahl von Räumen zur Verfügung. Es wurde dabei darauf geachtet, dass maximal zwei Befragungspaare in einem Raum parallel arbeiteten, um die Aufmerksamkeit der Kinder nicht zu sehr abzulenken. Alle nun folgenden Tests und Aufgaben wurden standardisiert instruiert. Die speziell geschulten Versuchsleiter achteten sehr darauf, dass die Kinder jede Aufgabenstellung richtig verstanden hatten. Falls es Verständnisprobleme bei einzelnen Aufgaben gab, wurden diese nochmals erläutert. Auch die Reihenfolge der Tests wurde standardisiert, wobei der Einsatzpunkt, also die erste Aufgabe, variiert wurde, falls zwei Erhebungen im gleichen Raum stattfanden. Dies geschah, um sicherzustellen, dass die Probanden nicht synchron den gleichen Test bearbeiteten und sich möglicherweise von den mitgehörten Antworten ihrer Klassenkameraden beeinflussen ließen. Insgesamt dauerte die Bearbeitung, abhängig von der Bearbeitungsgeschwindigkeit des Kindes, zwischen 50 und 80 Minuten. Zum Ende der Einzelbefragungen erhielten die Kinder einen Umschlag mit zwei Elternfragebögen (einen für Mütter, einen für Väter). Diese sollten die Kinder zu Hause abgeben, von ihren Eltern ausfüllen lassen, und im verschlossenen Umschlag wieder in die Schule mitbringen. Dort wurden die Daten dann entweder vom Erhebungsteam entgegengenommen oder von der jeweiligen Klassenlehrerin gesammelt und an die Universität geschickt. Außerdem konnte sich jedes Kind nach der Befragung eine kleine Aufmerksamkeit (Stifte, Bälle, etc.) als Dankeschön für die Teilnahme aussuchen. Der Instruktionsleitfaden, der die Reihenfolge der Tests wiedergibt, sowie die Einzelinstruktionen und Materialien für die Einzelbefragung und die Materialien für die Gruppenbefragungen finden sich im Anhang B.

III.5.2. Der Posttest T2

Der Untersuchungsablauf im Posttest gleicht bis auf wenige Ausnahmen dem des Prätests. Der Posttest startete am 16. und 17.01.2008 in der Grundschule Wandersleben und endete am 17. und 19.03.2008 in der Grundschule Unstruttal. Somit umfasst der Erhebungszeitraum für den Posttest ebenfalls eine Zeitspanne von neun Wochen, wobei in der Woche vom 03.02.2008 bis zum 10.02.2008 keine Erhebungen stattfanden, da in dieser Woche in Thüringen Schulferien (Winterferien) waren.

Die wenigen Veränderungen im Ablauf der Befragung entstanden durch das Weglassen beziehungsweise die Hinzunahme einzelner Instrumente. So wurden in der Gruppenbefragung der Kinder der gleiche Toleranzfragebogen und das Intended Behaviour Measure eingesetzt, nicht jedoch der Raven-Matrizen-Test. Stattdessen wurden einige Items zur Erfassung der sozialen Identifikation der Kinder eingefügt. Auch diese Items wurden standardisiert instruiert und alle Verständnisprobleme durch das Erhebungsteam ausgeräumt. Nach der Durchführung der Gruppenbefragung wurden die ausgeteilten Erhebungsmaterialien wieder vom Team codiert und eingesammelt. Außerdem wurden die Elternfragebögen den Kindern bereits im Anschluss an die Gruppenbefragung ausgeteilt. Durch die Herausnahme des Raven-Matrizen-Tests und die gestiegene Lesefähigkeit der Kinder konnte die Befragungsdauer der Gruppenbefragung auf 45 bis 50 Minuten verkürzt werden.

Bei der Einzelbefragung ergaben sich kleine Änderungen in der Reihenfolge der Tests. Zudem konnten einige Aufgaben vereinfacht und gekürzt werden (Chandler-Cartoons, Klassifikationstest, Moralinterview), während einige neue Aufgaben hinzukamen (soziale Identifikation, soziale Distanz, – s. Kapitel III.6. Messinstrumente). Durch die Änderungen konnte die Dauer der Einzelbefragung auf 45 bis 60 Minuten verkürzt werden. Die Reihenfolge der Tests, der Instruktionsleitfaden sowie die Erhebungsmaterialien des Posttests können im Anhang B betrachtet werden.

III.5.3. Das Follow-up T3

Die dritte Erhebungswelle fand nach den Sommerferien im Herbst 2008 statt. Sie begann am 08.09.2008 in der Grundschule Wandersleben und endete am 14.11.2008 in der Grundschule Hörselgau. Die Erhebungszeit wurde dabei von den zweiwöchigen Herbstferien unterbrochen und umfasste insgesamt einen Zeitraum von acht Wochen.

Der Ablauf der Untersuchung gleicht dabei bis auf wenige Ausnahmen dem Ablauf des Posttests. Zunächst fand die Gruppenbefragung mit dem Fragebogen und dem Intended Behavior Measure mit der teilnehmenden Klasse statt. Dabei wurde dem Intended Behavior Measure ein zusätzliches Szenario mit einem russischen Kind hinzugefügt. Ebenso wurde der Fragebogen in der Verkuytenskala um einige Items zu russischen Kindern ergänzt. Insgesamt konnte die Gruppenbefragung wie zum Posttest durchgeführt werden und dauerte etwa 45 Minuten.

Im Anschluss an die Gruppenbefragung erhielten die Kinder einen neuen Fragebogen für ihre Eltern, und es wurde – wie beim Posttest – die Einzelbefragung mit den teilnehmenden Kindern durchgeführt. Nach Beendigung der Einzelbefragung durften sich die Kinder auch hier ein kleines Dankeschön aussuchen. Die Einzelbefragung nahm pro Kind etwa 45 Minuten in Anspruch.

Die Reihenfolge der Tests, der Instruktionsleitfaden sowie die Erhebungsmaterialien des Follow-up können ebenfalls im Anhang B betrachtet werden. Die unten stehenden Tabellen 13 und 14 geben einen Überblick zur Anwendung und Reihenfolge der eingesetzten Instrumente der Befragungen von Kindern, Lehrern und Eltern.

Tabelle 13

Instrumente der Gruppen- und Einzelbefragung der Kinder nach Messzeitpunkten

Messzeitpunkt	Gruppenbefragung	Einzelbefragung
T1	demographische Angaben	Ambiguous Situation Task
	Toleranzskala	Klassifikationstest
	Interpersonale	Chandler-Cartoons
	Ambiguitätstoleranz	MRA
	Verkuytenskala	Moralinterview
	Similarity Task	Wissensquiz
	Raven Matrizentest	Soziales Problemlösen
	Intended Behavior Measure	Münztest
T2	demographische Angaben	Ambiguous Situation Task
	Toleranzskala	Klassifikation
	Interpersonale	Identifikation
	Ambiguitätstoleranz	soziale Distanz
	Verkuytenskala	Wissensquiz
	Similarity Task	Chandler-Cartoons
	soziale Identifikation	MRA
	Intended Behavior Measure	Moralinterview
T3	demographische Angaben	Ambiguous Situation Task
	Toleranzskala	Klassifikation
	Interpersonale	Identifikation
	Ambiguitätstoleranz	soziale Distanz
	Verkuytenskala	Wissensquiz
	Similarity Task	Chandler-Cartoons
	soziale Identifikation	MRA
	Intended Behavior Measure	Moralinterview
		Münztest
		Soziales Problemlösen

Anmerkungen. Die Instrumente wurden in der Reihenfolge von oben nach unten zu jedem Messzeitpunkt administriert. Zunächst die Gruppenbefragung, im Anschluss die Einzelbefragung.

Tabelle 14

Instrumente der Befragung der Eltern und Lehrer nach Messzeitpunkten

Messzeitpunkt	Elternbefragung	Lehrerbefragung
T1	Sozio-demographische Angaben Erziehungsstil Familienkohäsion SDQ Openness SDO Einstellungen gegenüber sozialen Gruppen	SDQ Items zu Vorurteilsneigung der Kinder
T2	Sozio-demographische Angaben SDQ SDO Einstellungen gegenüber sozialen Gruppen	SDQ Items zu Vorurteilsneigung der Kinder
T3	Sozio-demographische Angaben SDQ SDO Einstellungen gegenüber sozialen Gruppen	

Anmerkungen.

III.6. *Messinstrumente*

Nicht alle oben genannten Messungen waren für die Evaluation von PARTS von zentraler Bedeutung. Einige Messungen wurden auch zur Erforschung entwicklungspsychologischer Aspekte durchgeführt. Im Folgenden werden nur die für die Evaluation des Interventionsprogramms relevanten Instrumente vorgestellt. Dabei wird auf den Ursprung, die Inhalte, die Durchführung, die Auswertung und auf methodische Qualitätsmerkmale des jeweiligen Instrumentes (interne Konsistenzen) eingegangen. Eine vollständige Darstellung des verwendeten Materials zu den Messungen findet sich im Anhang B.

III.6.1. *Implementationsmaße*

Implementation – Protokollbogen zur Verhaltenseinschätzung

Ursprung. Die Protokollbögen zur Verhaltenseinschätzung der an der Intervention teilnehmenden Kinder basieren auf Implementationsbögen, die bereits von Beelmann (2000) bei anderen Evaluationsstudien erfolgreich eingesetzt wurden. In den hier genutzten Bögen werden die Beteiligung an den Interventionseinheiten (On-Task-Verhalten), das Störverhalten bzw. unaufmerksames Verhalten (Off-Task-Verhalten) und der Spaß der Kinder an der jeweiligen Interventionseinheit eingeschätzt.

Inhalt. Die Protokollbögen erfragen die oben genannten drei Bereiche des Verhaltens der Kinder während der Interventionssitzungen. Die Skala „On-Task-Verhalten“ erfasst, wie stark sich das Kind mit den Spielen und Übungen in der jeweiligen Programmeinheit beschäftigt. Dazu zählen sowohl korrekte als auch inkorrekte themenrelevante Beiträge, aber auch interessiert Zuhören und aufmerksames Verfolgen der Trainingsinhalte. Auf der Skala „Off-Task-Verhalten“ wird eingeschätzt, wie sehr sich das Kind während der Trainingseinheit mit anderen Dingen beschäftigt. Sei es störendes Verhalten oder auch nur desinteressiertes Verhalten, wie aus dem Fenster starren und seinen eigenen Gedanken nachhängen. Schließlich wird der Spaß des Kindes an der jeweiligen Übungseinheit eingeschätzt.

Durchführung. Nach bzw. während jeder Trainingseinheit füllen der anwesende Lehrer (meist der Klassenlehrer) sowie einer der Trainer unabhängig voneinander einen Protokollbogen zur Verhaltenseinschätzung aus. Auf diesen

Bögen sind die teilnehmenden Kinder der entsprechenden Klasse aufgeführt. Zu jedem Kind werden die drei Bereiche On-Task-Verhalten, Off-Task-Verhalten und Spaß eingeschätzt. Dabei geben Lehrer und Trainer bei On/Off-Task-Verhalten auf einer Skala von 0 = gar nicht bis 3 = sehr stark an, wie sie das Verhalten der einzelnen Kinder einschätzen. Der Spaß der Kinder an der Einheit wird von 0 = sehr wenig bis 2 = viel in einer dreigestuften Skala bewertet.

Auswertung. Somit ergeben sich für jeden Teilnehmer zu jeder Interventionseinheit die drei Werte: On-Task-Verhalten und Off-Task-Verhalten mit einem Wertebereich von 0-3 sowie Spaß mit einem Wertebereich von 0-2.

Methodische Qualität. Da zu jeder Einheit pro Teilnehmer und Subskala nur ein einziger Wert vorliegt, kann hier keine interne Konsistenz berichtet werden. Die Validität der Messung kann jedoch über die Beobachterübereinstimmung der ausfüllenden Lehrer und Trainer ermittelt werden. Die Ergebnisse hierzu werden ausführlich im Kapitel IV.3. Implementationsanalyse berichtet.

III.6.2. Kognitive und sozial-kognitive Konstrukte

Intelligenz – Raven Matrizen-Test

Ursprung. Die Standardform des Progressiven Matrizen-Tests wurde 1938 von Raven in seiner ersten Version veröffentlicht. Hier kam er in der deutschen Form von 1998 zum Einsatz (Heller, Kratzmeier & Lengfelder, 1998). Ziel des Tests ist es, die kognitiven Fähigkeiten (allgemeine Intelligenz, räumliche Wahrnehmungsdifferenzierung, Analogieschlussdenken, Regelerkennung und Prinzipienanwendung) zu messen.

Inhalt. Der Test besteht aus fünf Aufgabensets mit jeweils zwölf Aufgaben, von denen die ersten vier Aufgabensets A-D für die Testung von Grundschulern eingesetzt werden. Die Schwierigkeit und Komplexität der Aufgaben nimmt dabei von Set zu Set immer mehr zu. Die einzelnen Aufgaben bestehen aus vorgegebenen Strichzeichnungen oder einfachen Figuren, die nach bestimmten Prinzipien aufgebaut sind. Dabei ist immer ein Teil des Musters ausgelassen. Die Aufgabe des Probanden besteht darin, das Prinzip der jeweiligen Aufgabe herauszufinden und danach das fehlende Teil des Musters aus den Lösungsvorschlägen richtig auszuwählen. Für die vorliegende Studie wurden acht Matrizen, jeweils zwei aus den Sets A bis D, nach der Schwierigkeit und Trennschärfe der Items ausgewählt.

Durchführung. Der Test wurde im Rahmen der Gruppenbefragung durchgeführt. Den Kindern wurden Aufgabenblätter ausgeteilt. Zudem wurden die Aufgaben mit Hilfe eines Tageslicht-Projektors an die Wand projiziert, um die Instruktion zu vereinfachen. Die Kinder wurden instruiert, die passenden Teilstücke zur korrekten Ergänzung des Musters durch Anstreichen auf ihrem Bogen zu markieren.

Auswertung. Die einzelnen Items wurden mit 0 = „falsch“ und 1 = „richtig“ kodiert und zu einem Summenscore addiert. Somit ergibt sich ein Wertebereich von 0 bis 8.

Methodische Qualität. Die Reliabilität des Gesamttests (Sets A-D = 48 Items) wird für Grundschüler der zweiten Klasse mit $\alpha = .91$ (Heller, Kratzmeier & Lengfelder, 1998) angegeben. In der vorliegenden Untersuchung ist die Reliabilität mit einem $\alpha = .43$ deutlich geringer. Dies liegt ohne Zweifel an der, aus testökonomischen Gründen notwendigen, Kürzung der Skala und der Auswahl der Einzelitems aus unterschiedlichen Sets mit verschiedenen Schwierigkeitsstufen.

Multiple Klassifikationsfähigkeit – Klassifikationstest

Ursprung. Der Klassifikationstest beruht auf dem Multiple Classification Skill Task, der 1993 von Bigler und Liben eingesetzt wurde, und der die multiple Klassifikationsfähigkeit der Kinder erfassen soll.

Inhalt. Der Test besteht aus zwei Teilen: Der multiplen Klassifikation und der Reklassifikation. Die Probanden sollen zwei Sets von Stimulusmaterial, einmal Personen und einmal Kleidungsstücke, in ein Vierfelderschema sortieren. Das Stimulusmaterial besteht aus jeweils acht Kärtchen, auf denen Personen bzw. Kleidungsstücke abgebildet sind, die sich auf drei Dimensionen unterscheiden, ansonsten aber identisch sind. Diese Kärtchen sollen so sortiert werden, dass die Personen/Teile, die zusammengehören, also die sich ähnlich sind, zusammen in einem Feld liegen, während die Personen/Teile die sich weniger ähnlich sind in getrennten Feldern liegen. Die Probanden sollen das Stimulusmaterial also nach den drei Merkmalsdimensionen sortieren.

Durchführung. Dieser Test war Bestandteil der Einzelbefragung. Zunächst erhielten die Probanden acht Personenkärtchen, die sich auf den Dimensionen Geschlecht, Hautfarbe und Gewicht unterschieden. Die Kärtchen wurden, zufällig gemischt, vor den Probanden ausgebreitet. Daneben lag ein weißes DIN A4-Blatt mit

einem Kreuz in der Mitte. Dann wurden die Kinder instruiert, die Kärtchen so dem Vier-Felder-Schema zuzuordnen, dass am Ende die Kärtchen gemeinsam in einem Feld liegen, die zusammengehören.

Nachdem die Kinder die Kärtchen sortiert hatten, wurden sie nach der Begründung für ihre Sortierung gefragt. Der Versuchsleiter protokollierte auf dem Auswertungsbogen sowohl das Muster, nach welchem der Proband die Kärtchen sortiert hatte als auch dessen Begründung. Nun sollten die Kinder im nächsten Schritt das Material reklassifizieren. Sie wurden gefragt, ob sie noch eine andere Möglichkeit kannten, die Karten sinnvoll in das Gitter zu ordnen. Wenn ja, legten die Kinder ihre zweite Lösung und wurden wieder nach den Gründen für diese Sortierung gefragt. Dies wurde noch ein drittes Mal wiederholt. Falls ein Kind vor der dritten Lösung keine neue Idee mehr hatte, wurde ihm eine noch nicht vorgekommene Musterlösung vom Versuchsleiter vorgelegt. Das Kind sollte diese Lösung anschauen und begründen, warum die Kärtchen darauf so angeordnet wurden. Nachdem der Versuchsleiter alle Lösungen und Begründungen protokolliert hatte, wurde das gleiche Verfahren mit dem zweiten Stimulusmaterialset fortgeführt.

Das zweite Set bestand aus acht Kärtchen, auf denen jeweils ein Kleidungsstück abgebildet war. Die Kärtchen unterschieden sich auf den drei Dimensionen Farbe, Form (Kappe vs. Schuhe) und Emblem. Ansonsten entsprach der Ablauf dieser Testreihe genau der bereits oben geschilderten Reihe mit dem Personen-Stimulusmaterial. Dieses Stimulusset (Bekleidung) wurde ausschließlich zum ersten Messzeitpunkt durchgeführt, um eine Basisrate der multiplen Klassifikationsfähigkeit mit nicht-sozialen Stimuli zu erfassen. Für die Evaluation von PARTS wurde ausschließlich die Serie mit sozialen Stimuli verwendet, da diese die inhaltlich relevanten Bereiche aus der Intervention erfasst.

Auswertung. Beide Testserien wurden mit jeweils einem Punkt für eine richtig sortierte Dimension und einem Punkt für eine richtige Begründung der gelegten Dimension bewertet. Somit ergibt sich für jede Serie eine Gesamtpunktzahl von 0 bis 12.

Methodische Qualität. Bigler und Liben (1993) geben einen zufriedenstellenden Reliabilitätskoeffizienten von $\alpha = .70$ für den Test in ihrer Untersuchung an. In der vorliegenden Arbeit ist die interne Konsistenz des Tests mit den sozialen Stimuli zum ersten Messzeitpunkt $\alpha = .63$, zum zweiten Messzeitpunkt bei $\alpha = .75$ und bei der dritten Messung $\alpha = .60$. Die Reliabilität für den

Klassifikationstest mit den nicht-sozialen Stimuli im ersten Messzeitpunkt ergibt einen Wert von $\alpha = .85$. Insgesamt also auch eine zufriedenstellende interne Konsistenz.

Interkulturelles Wissen – Wissenstest

Ursprung. Der Wissenstest wurde eigens für das Projekt „Förderung von Intergruppenbeziehungen bei Kindern“ entwickelt. Er dient dazu, interkulturelles Wissen von Kindern zu messen.

Inhalt. Der Test beinhaltet 15 Items die unterschiedliche Wissensbereiche erfragen. Die ersten sieben Fragen beziehen sich dabei auf allgemeines Wissen über Geographie (wie zum Beispiel „Welche Kontinente gibt es auf unserer Erde?“). Die anderen acht Items wurden auf die Inhalte der Wissensintervention von PARTS abgestimmt und befassen sich mit interkulturellem Wissen, wie z.B. Gemeinsamkeiten und Unterschieden zwischen Menschen verschiedener Kulturen, geographischen Entwicklungsbedingungen für unterschiedliche Kulturen, Gründen für das Zusammentreffen von Menschen verschiedener Kulturen sowie Metawissen über kulturelle Begegnungen (z.B. „Weißt Du was ein Vorurteil ist? Erkläre bitte den Begriff!“). Für die Evaluation von PARTS wurde ausschließlich die Subskala des interkulturellen Wissens verwendet, da diese die inhaltlich relevanten Bereiche aus der Intervention erfasst.

Durchführung. Die Fragen wurden den Probanden im Rahmen der Einzelbefragung vorgelesen und die Antworten von dem jeweiligen Versuchsleiter auf dem Antwortbogen notiert. Falls die Kinder Verständnisschwierigkeiten bei einzelnen Fragen hatten, wurden diese vom Versuchsleiter ausgeräumt.

Auswertung. Die Items haben unterschiedliche Antwortformate. Manche Fragen sind offen gestellt, andere bieten Multiple-Choice-Antwortmöglichkeiten und wieder andere erfordern die Nennungen mehrerer richtiger Lösungen. Jede richtige Antwort wird dabei mit Punkten honoriert, die aufsummiert den Gesamtwert im Wissenstest ergeben. Dabei wurden die Items, welche die Aufzählung mehrerer richtiger Antworten beinhalteten, gewichtet, damit nicht einige wenige Items zu stark in den Gesamtwert eingehen. Der Wertebereich für den Gesamtwert der Skala interkulturellen Wissens liegt zwischen 0 und maximal 24 Punkten.

Methodische Qualität. Die Reliabilität der Skala interkulturellen Wissens beläuft sich zu t1 auf $\alpha = .56$, zu t2 auf $\alpha = .64$ und zu t3 auf $\alpha = .71$ und ist somit als zufriedenstellend einzuschätzen.

Perspektivenübernahme – Chandler Cartoons

Ursprung. Der Chandler Task wurde von Michael Chandler entwickelt (1973) und basiert auf einer Idee von Flavell, Botkin, Fry, Wright und Jarvis (1968). Bei dieser Aufgabe geht es darum, die kognitive Perspektivenübernahmefähigkeit von Kindern zu messen.

Inhalt. Dazu werden den Kindern kurze Bildergeschichten vorgelegt, die eine psychologisch-kausale Handlungskette beinhalten. In der Flugzeug-Geschichte begleitet beispielsweise ein kleines Mädchen ihren Vater zum Flughafen und beobachtet traurig wie er davonfliegt. Später fängt das kleine Mädchen (die Protagonistin der Geschichte) an zu weinen, als es ein Spielzeugflugzeug geschenkt bekommt, das dem ähnelt, in dem ihr Vater abgereist ist. Dies beobachtet ein später auftauchender Postbote (der Bystander dieser Geschichte).

Nun wird anhand von Nacherzählungen geprüft, ob die Kinder die Informationen, die nur ihnen zur Verfügung stehen (z.B. dass der Vater mit dem Flugzeug verreist ist) adäquat von denen trennen können, über die auch der Bystander (der nur das Weinen des Mädchens bei Erhalt des Spielzeugflugzeugs beobachtet hat) verfügt.

Durchführung. Die Probanden sollten im Rahmen der Einzelbefragung zunächst eine Bildergeschichte nacherzählen, die ihnen der Versuchsleiter Bild für Bild vorlegte. Falls aus dieser Nacherzählung deutlich wurde, dass die Teilnehmer die Geschichte nicht verstanden hatten, berichtete der Versuchsleiter sie und stellte sicher, dass der Handlungsablauf klar war. Im Anschluss daran sollten sie die Geschichte noch einmal, aus der Sicht eines nicht informierten Bystanders, der zu einem späteren Zeitpunkt in der Geschichte auftaucht, erzählen. Dabei wurden die Probanden instruiert, mit dem Bild zu beginnen, in dem der Bystander das erste Mal in der Geschichte auftaucht. Die Probanden wurden aufgefordert sich in die Rolle dieses Bystanders zu versetzen, und dabei besonders auf die Gedanken und Gefühle der in der Geschichte beteiligten Personen einzugehen. Falls die Probanden nicht von sich aus auf die Gedanken und Gefühle der Personen eingingen, wurden diese vom Versuchsleiter gesondert erfragt, nachdem die Teilnehmer mit ihren Erzählungen fertig waren. Alles was die Probanden zu den beiden Aufgaben erzählten, wurde parallel vom Versuchsleiter protokolliert.

Für die vorliegende Arbeit wurden die beiden Cartoons Sandburg und Flugzeug ausgewählt. Um die ohnehin umfangreiche Testbatterie und die kognitive

Belastung der Kinder zu mindern, und um die standardisierte Anwendung des Messinstruments zu erhöhen, wurde zum zweiten Messzeitpunkt die Prozedur des Chandler Task leicht abgewandelt. Die Probanden mussten hier nicht zuerst die komplette Geschichte erzählen, bevor sie den Bystander-Bericht abgeben sollten. Die komplette Geschichte wurde dieses Mal vom Versuchsleiter parallel zur Präsentation der Bilder vorgelesen, und die Probanden konnten sich auf den Bericht des Bystanders konzentrieren. Nichts desto trotz bestand die Aufgabe in der Übernahme der Rolle des Bystanders und der Trennung der Informationen, die nur der Proband besaß, vom Informationsstand des Bystanders.

Auswertung. In der ursprünglichen Fassung des Tests wurden die Bystanderberichte der Probanden fünfstufig kodiert. Da jedoch bei der Auswertung der Daten eine Kategorie nicht vorkommt, konnte die Skala auf vier Stufen reduziert werden. Je deutlicher die Bystanderberichte Intrusionen in Form von privilegiertem Wissen enthalten, desto höher ist die Punktzahl. Berichte, in denen explizit Informationen verwendet wurden, welche nur dem Probanden, nicht aber dem Bystander zur Verfügung standen, wurden mit vier Punkten kodiert. Drei Punkte wurden vergeben, wenn die Intrusionen durch probabilistische Abschwächungen entschärft wurden (z.B. „Der Postbote denkt, dass ihr Vater *vermutlich* mit dem Flugzeug verreist ist.“). Ein Wert von zwei wurde vergeben, wenn die Intrusion in einer Reihe von anderen, nicht egozentrischen Alternativen vorgetragen wurde (z.B. „Der Postbote denkt, dass ihr das Geschenk nicht gefällt, oder dass Jemand Wichtiges mit dem Flugzeug verreist ist, oder so.“). Einen Punkt gab es schließlich für Berichte, in denen klar zwischen Informationen, die dem Bystander zur Verfügung standen und Wissen, das nur dem Probanden selbst zur Verfügung stand, getrennt wurde.

Die beiden Werte aus den jeweiligen Szenarien wurden für die weiteren Auswertungen zu einem Mittelwert pro Messzeitpunkt zusammengefasst. Der Wertebereich reicht von 1 bis 4 wobei ein hoher Wert für viele Intrusionen und damit für eine geringere Ausprägung der Perspektivenübernahmefähigkeit spricht. Aus Gründen der Interpretierbarkeit der Daten wurden die Mittelwerte umkodiert, so dass in der endgültigen Analyse hohe Werte ein hohes Niveau an Perspektivenübernahmefähigkeit widerspiegeln.

Methodische Qualität. In der ursprünglichen Fassung des Tests von Chandler wurden zehn verschiedene Cartoons verwendet, die eine Split-half Reliabilität von

0.92 aufwiesen (Chandler, 1973). Auch die Interrater-Reliabilität bewegte sich mit 0.94 in einem ähnlich hohen Bereich. In späteren Studien (z.B. Chandler, Greenspan & Barenboim, 1974) wurde der Chandler Task in einer gekürzten Fassung mit nur drei Cartoons eingesetzt. Die Reliabilitäten der Chandler-Cartoons in der vorliegenden Arbeit erweisen sich, ermittelt über Cronbachs α , allerdings als problematisch. Zum ersten Messzeitpunkt hat die Skala eine Reliabilität von $\alpha = .28$. Zu den Messzeitpunkten zwei und drei konnte die Zuverlässigkeit auf ein $\alpha = .45$ gesteigert werden. Eine ähnlich geringe interne Konsistenz der Chandler-Skala berichtet Rubin (1978). Er nutzte vier Chandler-Cartoons und erhielt durchschnittliche Interitemkorrelationen von $r = .33$ bei einer Stichprobe von Drittklässlern. Allerdings weisen andere Maße zur Messung von Perspektivenübernahme bei Kindern ähnlich geringe interne Konsistenzen auf (Rubin, 1978, Kurdek, 1977).

Zwischen der Chandler-Skala und einigen Vorurteilsmaßen existieren allerdings systematische Zusammenhänge (Steudel, 2010). Dies kann ein Hinweis für eine noch relativ valide Messung des Konstrukts „Perspektivenübernahmefähigkeit“ durch die Chandler-Skala sein.

Soziale Informationsverarbeitung – Test zum sozialen Problemlösen

Ursprung. Der Test zum sozialen Problemlösen wurde in ähnlicher Form bereits von Beelmann (2000) bei Vorschulkindern eingesetzt und basiert auf einer Idee von Döpfner und Kollegen (PIPS-Test, 1989). Die hier vorliegende Fassung wurde an die Altersstufe der teilnehmenden Kinder angepasst. Die Fähigkeit der Kinder zur Lösung von sozialen Problemen wird anhand von ambigen sozialen Problemszenarien erfasst.

Inhalt. Den Kindern werden alltagsnahe Szenarien vorgelesen, die eine soziale Interaktion mit ambigem konfliktreichem Inhalt beschreiben. Daraufhin werden die Kinder nach ihrer Interpretation dieser Szenarien und ihrer Reaktion auf diese gefragt.

Durchführung. Der Test zum sozialen Problemlösen fand in der Einzelbefragung statt. Drei Szenarien mit unterschiedlichen Inhalten wurden den Kindern vorgelesen. Im Anschluss an jedes der Szenarien sollten die Kinder beantworten, warum es ihrer Meinung nach zu dieser Situation gekommen sei, und was sie in einer derartigen Situation tun würden. Die Antworten der Kinder wurden

wörtlich auf dem Protokollbogen notiert. Ebenso wurde mit den beiden weiteren Szenarien verfahren.

Auswertung. Bei der Auswertung des Tests wurden die Antworten der Kinder verschiedenen Kategorien zugeordnet. Die Frage nach der Ursache für die Situation wurde in zwei Kategorien eingeteilt: freundlich/neutrale Erklärung (mit 2 kodiert) versus feindliche Erklärung (mit 1 kodiert). Die Antworten der Kinder auf die Frage nach ihren Verhaltensweisen in einer derartigen Situation wurden als kompetent (mit 3 kodiert), unsicher (mit 2 kodiert) oder aggressiv (mit 1 kodiert) kategorisiert. Die sechs sich insgesamt ergebenden Werte wurden für die weitere Auswertung addiert, so dass sich eine Skala mit einem Wertebereich von 6-15 ergibt, wobei hohe Werte eine höhere soziale Kompetenz und günstigere soziale Informationsverarbeitung anzeigen.

Methodische Qualität. Die vorliegenden Daten zu der oben beschriebenen Skala sozialer Informationsverarbeitung weisen nur eine geringe Reliabilität von $\alpha = .43$ zum ersten Messzeitpunkt und $\alpha = .41$ zum zweiten bzw. $\alpha = .42$ zum dritten Messzeitpunkt auf. Dies ist vermutlich der hohen inhaltlichen Eigenvarianz der einzelnen Szenarien zuzuschreiben. Eine Reduzierung der Skala erbringt allerdings keine Verbesserung der internen Konsistenz. Daher wird die Skala zur sozialen Informationsverarbeitung in oben beschriebener Weise verwendet.

III.6.3. Gruppenbezogene soziale Einstellungen

Gruppenbezogene soziale Einstellungen können nach unterschiedlichen Merkmalen eingeteilt werden. So können diese Einstellungen nach dem Grad der Abstraktheit des Zielstimulus unterschieden werden. Manche Einstellungsmaße beziehen sich auf Einstellungen gegenüber einer gesamten Gruppe, während andere Einstellungsmaße ein möglichst prototypisches Mitglied der jeweiligen Gruppe als Zielstimulus nutzen. Man kann dementsprechend von gruppenzentrierten versus personenzentrierten Einstellungsmaßen sprechen.

Weiterhin besteht die Möglichkeit, die Einstellungen gegenüber einer Fremdgruppe einzeln zu betrachten oder diese Einstellung mit der Einstellung gegenüber der Eigengruppe zu referenzieren. Diese sogenannten Biasmaße, oder Einstellungsverzerrungen, werden über die Differenz zwischen der Eigengruppeneinstellung und der Fremdgruppeneinstellung gebildet (Allport, 1954; Brewer, 1999, s. Kapitel II.1. Grundlagen). Sie haben den Vorteil, generelle Antworttendenzen des einzelnen Probanden zu berücksichtigen und die Einstellungen zur Fremdgruppe, im Vergleich zur Einstellung zur Eigengruppe in einem Kennwert darzustellen. Ein solches Vorgehen ist auch deshalb sinnvoll, weil direkte negative Äußerungen gegenüber anderen in dieser Altersgruppe relativ selten sind (Nesdale, 2004). Für die Auswertungen bei PARTS werden bei gruppenbezogenen sozialen Einstellungen daher ausschließlich Biaswerte verwendet.

Schließlich können gruppenbezogene soziale Einstellungen noch nach der angesprochenen Erlebensebene unterschieden werden. So gibt es Einstellungen mit kognitivem, mit affektivem oder konativem Schwerpunkt.

Im Folgenden werden die für die vorliegende Evaluationsstudie relevanten Einstellungsmaße nach den oben beschriebenen Merkmalen eingeordnet und vorgestellt.

III.6.4. Gruppenbezogene soziale Einstellungen mit kognitivem Schwerpunkt

Unter gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt versteht man, wie bei Stereotypen, hauptsächlich Attributzuordnungen, also die Zuschreibung von bestimmten Eigenschaften zu einer Personengruppe. Diese Eigenschaften können eine positive, eine negative oder eine neutrale Valenz aufweisen.

Biasmaße via Attributzuordnung (personenorientiert)

– Multi-response Racial Attitude Measure

Ursprung. Das Multi-response Racial Attitude Measure (MRA) wurde 1995 von Anna Beth Doyle und Frances Aboud entwickelt. Der Test dient dazu, ethnische Vorurteile bei Kindern zu messen. Er basiert auf dem Preschool Racial Attitude Measure II (PRAM II; Williams, Best, Boswell, Mattson & Graves, 1975) und reiht sich damit in die Tradition der Messung von Vorurteilen via Eigenschaftszuschreibungen ein, die bereits 1938 von Horowitz und Horowitz aufgegriffen wurde.

Inhalt. Beim MRA werden den Versuchspersonen drei Zeichnungen von gleichgeschlechtlichen Kindern unterschiedlicher Ethnie vorgelegt. Die Probanden weisen dann diesen abgebildeten Kindern positive und negative Eigenschaften in Form von drei Kärtchen pro Eigenschaft zu. Aus der Anzahl der zugewiesenen positiven und negativen Eigenschaften kann dann auf die ethnienpezifische Einstellung der Probanden geschlossen werden. Dieses Vorgehen von Doyle und Aboud wurde in der vorliegenden Untersuchung übernommen, allerdings wurden die zu beurteilenden Ethnien an in Deutschland lebende Minoritäten angepasst. Statt der ursprünglichen weißen, schwarzen und indianischen Kinder, wurden hier deutsche, russische und chinesische Kinder eingeschätzt. Außerdem hatten die Teilnehmer in dieser Untersuchung explizit die Möglichkeit, Eigenschaften keinem der Kinder zuzuordnen. Somit sollte die forced-choice-Problematik der ursprünglichen Instruktion von Doyle und Aboud umgangen werden, bei der die Eigenschaften entweder einer oder allen Ethnien zugeordnet werden mussten.

Durchführung. Den Probanden wurden im Rahmen der Einzelbefragung zunächst drei Zeichnungen von Kindern vorgelegt. Die Zeichnungen stellten ein deutsches, ein russisches und ein asiatisches Kind dar. Den männlichen

Versuchsteilnehmern wurden Bilder von Jungen vorgelegt, den weiblichen Versuchsteilnehmern Bilder von Mädchen. Die Zeichnungen zeigten Kopf und Schulterpartie der Kinder, wobei die spezifische Ethnie über die Haut- bzw. Haarfärbung, die Gesichtszüge sowie die zugehörigen Namen deutlich gemacht wurde. Jedem Bild wurde ein leeres Kästchen zugeordnet, in das die Kinder später die Eigenschaftskärtchen legen sollten. Der Versuchsleiter stellte die abgebildeten Kinder mit Namen und Herkunftsländern vor und erklärte die Aufgabe.

Jedem Kind wurden jeweils drei Kärtchen vorgelegt, die eine bestimmte Eigenschaft repräsentierten. Daraufhin wurde das Kind gebeten, diese Kärtchen zu dem Bild der Kinder zu legen, zu denen die entsprechende Eigenschaft passe. Dabei konnten sie die Karten auf alle Kinder verteilen, aber auch einem Kind mehrere Karten zuweisen oder Karten an den Versuchsleiter zurückgeben, wenn sie der Meinung waren, nicht alle verteilen zu müssen. Nachdem die Kinder eine Einführungsaufgabe bearbeitet hatten, wurden sie gefragt, warum sie die Karten entsprechend verteilt hatten, um zu prüfen, ob die Kinder den Ablauf der Aufgabe korrekt verstanden hatten. Wenn alle Fragen geklärt waren, legte der Versuchsleiter die eigentlichen Testitems vor. Der Versuchsleiter protokollierte jeweils die Verteilung und ging zum nächsten Item über. Es wurden insgesamt 16 Items bearbeitet. Sieben Items waren positiv (lustig, gute Ideen, kann man trauen, ehrlich, freundlich, liebenswürdig, hilfsbereit), sieben Items negativ (unfreundlich, gemein, ärgert andere, Angeber, egoistisch, prügelt sich, mag nicht (mit anderen) spielen) und zwei Items neutral (mag Sport, mag Singen). Dabei wurden die positiven, negativen und neutralen Items in einer standardisierten gemischten Reihenfolge präsentiert. Zum Abschluss sollten die Kinder noch auf das Bild desjenigen Kindes zeigen das ihnen am ähnlichsten sei.

Auswertung. Bei der Auswertung des Tests wurden sechs verschiedene Werte berechnet. Ein positiver und ein negativer Wert für jede beurteilte Ethnie. Dabei zählte jedes zugeordnete positive oder negative Eigenschaftskärtchen als ein Punkt (vgl. Black-Gutman & Hickson, 1996). Somit ergibt sich für jeden Wert ein Wertebereich von 0 – 21. Da es sich beim MRA um ein vergleichendes Einstellungsmaß handelt, werden die positiven und negativen Werte der einzelnen Ethnien zueinander in Beziehung gesetzt, so dass sich zwei Biaswerte pro Fremdgruppe berechnen lassen. So kann ein Maß für die Eigengruppenpräferenz im Vergleich zu einer Fremdgruppe gebildet werden, indem die Anzahl der positiven

Zuweisungen zur Fremdgruppe von der Anzahl der positiven Zuweisungen zur deutschen Gruppe abgezogen werden. Somit ergeben sich zwei Eigengruppenpräferenzwerte, einmal im Vergleich zwischen dem deutschen und dem chinesischen Kind, und einmal im Vergleich zwischen dem deutschen und dem russischen Kind. Diese Skalen verfügen über einen Wertebereich der von -21 (maximale Bevorzugung des Fremdgruppenkindes) bis 21 (maximale Bevorzugung des Eigengruppenkindes) reicht. Ähnlich lässt sich auch ein Maß für Fremdgruppenabwertung operationalisieren. Hierzu werden die Anzahl negativer Zuweisungen zum deutschen Kind von den negativen Zuweisungen zum Fremdgruppenkind abgezogen. Auch hier ergibt sich ein Wertebereich von -21 (maximale Abwertung des Eigengruppenkindes) bis 21 (maximale Abwertung des Fremdgruppenkindes). Dabei korrelieren die Eigengruppenpräferenzwerte mit den Fremdgruppenabwertungswerten gegenüber der gleichen Fremdgruppe innerhalb eines Messzeitpunktes relativ hoch miteinander ($r = .71$ bis $r = .77$).

Methodische Qualität. Die internen Konsistenzen der einzelnen Skalen sind sehr gut. Sie liegen zum ersten Messzeitpunkt (MZP) zwischen $\alpha = .77$ (russische Kinder positiv) und $\alpha = .89$ (russische Kinder negativ). Zum zweiten MZP ergeben sich interne Konsistenzen zwischen $\alpha = .80$ (russische Kinder positiv) und $\alpha = 0.89$ (russische Kinder negativ). Auch zum dritten Messzeitpunkt bleiben die internen Konsistenzen der einzelnen MRA-Skalen auf einem ähnlich hohen Level. Die Untergrenze von Cronbachs Alpha liegt hier bei $\alpha = .80$ (deutsche Kinder positiv) und die Obergrenze liegt bei $\alpha = .89$ (russische Kinder negativ). Doyle und Aboud (1995) berichten für die interne Konsistenzen ähnlich hohe Alphawerte zwischen $\alpha = .79$ und $\alpha = .91$ sowie eine Test-Retest Reliabilität bei Viertklässlern über einen Zeitraum von vier Monaten von .56 bis .82.

Biasmaße via Attributzuordnung (gruppe nzentriert) – Verkuyten-Skala

Ursprung. Den Kindern wurden im Fragebogen sechs Items zur Zuschreibung von Eigenschaften zu asiatischen bzw. deutschen Kindern vorgelegt. Diese Items basieren auf einer Skala, die Kinket und Verkuyten (1999) entwickelten und die von Feddes (Feddes, Noack & Rutland, 2009) für deutsche Kinder angepasst und genutzt wurde. Feddes' Version der Skala wurde in der vorliegenden Studie eingesetzt, wobei die Fremdgruppe in Anpassung an den thüringischen Erhebungskontext geändert wurde. Während Feddes Einstellungen gegenüber

türkischen Kindern erhob, wurde in der vorliegenden Studie nach den Eigenschaften von chinesischen Kindern gefragt. Zum dritten Messzeitpunkt wurde die Skala um entsprechende Items zu russischen Kindern erweitert.

Durchführung. Die Kinder sollten im Fragebogen der Gruppenbefragung angeben, wie viele deutsche beziehungsweise chinesische Kinder *freundlich*, *höflich* und *klug* seien. Dabei hatten sie die vier Antwortmöglichkeiten *keine*, *manche*, *viele* oder *alle*. Zusätzlich wurden sie direkt gefragt, wie sehr sie deutsche bzw. chinesische und russische Kinder mögen. Auch hier hatten sie ein viergestuftes Antwortformat von 1 = „gar nicht“ bis 4 = „sehr“.

Auswertung. Somit ergeben sich drei vier-Item-Skalen zur Bewertung der Gruppe der deutschen sowie der chinesischen und zusätzlich zum dritten Messzeitpunkt der russischen Kinder. Um individuelle Antworttendenzen zu berücksichtigen, wurde bei diesen Skalen ein Biaswert berechnet, indem die Einstellungen gegenüber den Fremdgruppen (chinesische Kinder und russische Kinder) von den Werten der Eigengruppe (deutsche Kinder) abgezogen wurden. Somit ergeben sich Biaswerte für die jeweilige Fremdgruppe, wobei positive Werte eine Bevorzugung der Eigengruppe angeben, und negative Werte eine Bevorzugung der Fremdgruppe indizieren. Der Wertebereich der Biaswerte liegt zwischen +3 (starke Bevorzugung der Eigengruppe) und -3 (starke Bevorzugung der Fremdgruppe).

Methodische Qualität. Die Skalen hatten im Original bei Kinket und Verkuyten (1999) noch ein viertes Item (*quarrelsome* - *streitsüchtig*) und waren fünfstufig skaliert. Kinket und Verkuyten gaben für ihre Skalen interne Konsistenzen von $\alpha = .61$ für die Eigengruppen-Skala und $\alpha = .68$ für die Fremdgruppen-Skala an. Derweil berichtet Feddes (2007) für die gekürzte Skala (drei Items) mit dem oben berichteten vierstufigen Antwortformat eine Reliabilität von $\alpha = .84$ sowohl für die Eigengruppen- als auch die Fremdgruppen-Skala bei deutschen Probanden. In der vorliegenden Studie erreichen die internen Konsistenzen Werte zwischen $\alpha = .56$ (Einstellung zu chinesischen Kindern zu t1) bis $\alpha = .73$ (Einstellungen zu russischen Kindern zu t3) und sind somit noch zufriedenstellend.

III.6.5. Gruppenbezogene soziale Einstellungen mit affektivem Schwerpunkt

Gruppenbezogene soziale Einstellungen mit affektivem Schwerpunkt bilden hauptsächlich Gefühle wie Angst, Sympathie oder das Nähe-Distanz-Bedürfnis gegenüber Mitgliedern einer bestimmten Gruppe ab. Sie können somit als eine Operationalisierung von Vorurteilen verstanden werden (s. Kapitel II.1. Grundlagen). Auch hier können die Zielstimuli gruppenzentriert oder personenzentriert dargestellt werden.

Biasmaße via Sympathieeinschätzungen

– gruppenzentriert und personenzentriert

Ursprung. Die Biasmaße via Sympathieeinschätzungen werden in der gruppenzentrierten Form aus einzelnen Items der Verkuyten-Skala (s.o.) und in der personenzentrierten Form aus einzelnen Items des Intended Behavior Measures (s.u.) gebildet. Im Kern steht die Frage, wie sehr eine Fremdgruppe oder eine Person, die Mitglied einer Fremdgruppe ist, im Vergleich zur Eigengruppe bzw. zum Eigengruppenmitglied gemocht wird. Somit kann dieser Biaswert als affektives Einstellungsmaß eingeordnet werden.

Inhalt und Durchführung. Die Items zur gruppenzentrierten Sympathieeinschätzung finden sich im Fragebogen, der in der Gruppenbefragung der Kinder eingesetzt wurde. Hier wurden die Kinder direkt gefragt „Wie sehr magst Du deutsche Kinder?“ bzw. „Wie sehr magst Du chinesische Kinder?“ und ab dem dritten Messzeitpunkt auch „Wie sehr magst Du russische Kinder?“. Die Teilnehmer hatten vier Antwortmöglichkeiten: „überhaupt nicht“, „eher nicht“, „eher schon“ oder „sehr“. Diese wurden mit traurigen und fröhlichen Smilies über den Antwortmöglichkeiten versehen, um die inhaltliche Valenz der Antwortkategorien für die Kinder deutlicher zu machen.

Bei der personenzentrierten Variante der Sympathieeinschätzung wurden die Kinder nach dem Verlesen des IBM-Szenarios im Rahmen der Gruppenbefragung (s.u.), gefragt: „Wie sehr würdest Du Renee¹⁸ mögen?“, bzw. „Wie sehr würdest Du Lee mögen?“, und ab dem dritten Messzeitpunkt auch „Wie sehr würdest Du Vladimir/Svetlana mögen?“. Auch hier sollten die Kinder anhand einer Smily-Skala

¹⁸ Renee wurde als deutsches Kind vorgestellt, Lee als chinesisches Kind und Vladimir bzw. Svetlana als russisches Kind.

angeben, wie sehr sie die entsprechende Person mögen würden. Allerdings war die Skala hier, wie bei allen anderen IBM-Items auch, fünfstufig.

Auswertung. Für die gruppenzentrierte Variante der Sympathieeinschätzung ergibt sich ein Wertebereich von 1 = überhaupt nicht bis 4 = sehr. Bei der personenzentrierten Version der Sympathieeinschätzung ergibt sich ein Wertebereich von 1 = überhaupt nicht bis 5 = sehr gern. In beiden Varianten wurden die Sympathiebekundungen gegenüber der Fremdgruppe/ den Fremdgruppenmitgliedern mit den Sympathiebekundungen gegenüber der Eigengruppe/ des Eigengruppenmitglieds in Relation gesetzt. Dazu wurden die Bewertungen der chinesischen und russischen Fremdgruppe von den Bewertungen der Deutschen abgezogen. Somit ergeben sich für die ersten beiden Messzeitpunkte ein Sympathiebiaswert gegenüber der Gruppe der chinesischen Kinder und ein Sympathiebiaswert gegenüber einem Mitglied der Gruppe der chinesischen Kinder. Zum dritten Messzeitpunkt kommt zu diesen beiden Werten das Pendant in Bezug auf russische Kinder hinzu.

Methodische Qualität. Da hier nur einzelne Items zur Einschätzung der Sympathie gegenüber den Gruppen und Gruppenmitgliedern verwendet wurden, lässt sich keine interne Konsistenz berechnen. Allerdings zeigen sich substantielle Zusammenhänge zwischen den beiden Sympathieeinschätzungsvarianten (Korrelationen von $r = .14$ bis $r = .47$). Dies spricht für eine relativ valide Messung des Konstrukts.

Biasmaß via soziale Distanz (gruppenzentriert) – Haus-Aufgabe

Ursprung. Die Haus-Aufgabe entspricht der „street exercise“ von Griffith und Nesdale (2006), die auf einer Idee von Valk und Karu (2001) beruht. Diese wurde ins Deutsche übersetzt und auf andere Fremdgruppen ausgerichtet.

Inhalt. Die Kinder werden gebeten, Familien, die der Eigengruppe oder bestimmten Fremdgruppen zugeordnet werden, in einer fiktiven Nachbarschaft zu platzieren. Je nachdem, wie nah oder weit weg vom „Haus des Probanden“ die Familien „einquartiert“ werden, ergibt sich eine geringere oder größere soziale Distanz. Diese kann als Indikator für die impliziten, eher emotionalen Einstellungen der Probanden herangezogen werden.

Durchführung: Den Probanden wurde im Rahmen der Einzelbefragung eine Zeichnung vorgelegt, die eine Reihe von neun einfachen Häusern darstellt. Die

Häuser bestanden aus einem Quadrat mit einem Spitzdach. Das Haus in der Mitte war grau schattiert und wurde dem Probanden als „Dein Haus“ vorgestellt. Nun sollte der Proband entscheiden, in welche der freien Häuser drei neue Familien einziehen sollten. Zielgruppe war dabei eine deutsche, eine asiatische und eine russische Familie. Der Versuchsleiter vermerkte die Antworten der Versuchsperson auf dem Protokollbogen.

Auswertung. Die Kodierung der Angaben der Kinder erfolgte von 1 = „Haus direkt nebenan“ bis 4 = „Haus am äußersten Rand“. Je größer die Zahl, desto größer die Distanz zum eigenen Haus und somit auch desto größer die soziale Distanz zu der jeweiligen Gruppe (Nationalität/Ethnie). Auch hier wurden Biaswerte berechnet, um die allgemeinen Distanztendenzen der Probanden herauszurechnen, indem die Distanzwerte der Eigengruppenfamilie von den Distanzwerten der Fremdgruppenfamilie abgezogen wurden. Somit ergeben sich ein Biaswert für die russische Familie und ein Biaswert für die asiatische Familie die jeweils einen Wertebereich von -3 (pro Fremdgruppenfamilie) bis +3 (pro Eigengruppenfamilie) aufweisen. Die Haus-Aufgabe wurde sowohl zum Posttest als auch zum Follow-up eingesetzt, jedoch nicht zum Prätest.

Methodische Qualität. Da für jede Ethnie nur ein einziges Item zur Messung der sozialen Distanz eingesetzt wurde, lässt sich keine interne Konsistenz des Instruments angeben.

III.6.6. Gruppenbezogene soziale Einstellungen mit konativem Schwerpunkt

Gruppenbezogene soziale Einstellungen mit konativem Schwerpunkt zeichnen sich dadurch aus, das Verhalten, oder zumindest Verhaltensintentionen gegenüber Mitgliedern einer Gruppe zu erfragen. Dies kann als Vorstufe zur Messung von diskriminierendem Verhalten verstanden werden.

Kontaktintentionen gegenüber Mitgliedern von Fremdgruppen (personenzentriert) – Intended Behavior Measure

Ursprung. Das Intended Behavior Measure (IBM) wurde 2006 von Cameron, Rutland, Brown und Douch entwickelt und eingesetzt. Dabei sollen die Kinder in der für die vorliegende Arbeit verwendeten Version angeben, wie sie sich in zwei hypothetischen Szenarien gegenüber einem deutschen Kind und einem chinesischen Flüchtlingskind verhalten würden (in der ursprünglichen Version waren es ein englisches- und ein Flüchtlingskind).

Inhalt und Durchführung. Die beiden, im Rahmen der Gruppenbefragung dargebotenen Szenarien sind identisch, außer, dass die Kinder sich im ersten Szenario ein gleichgeschlechtliches Flüchtlingskind namens Lee vorstellen und im zweiten Szenario ein gleichgeschlechtliches deutsches Kind namens Renee. Die Probanden sollen sich vorstellen diese Kinder im Park zu treffen und wurden gefragt, wie sie sich diesen gegenüber verhalten würden. Zu jedem Szenario wurden die Probanden befragt, wie gerne sie mit dem Kind spielen würden, wie sehr sie es mögen würden, wie gern sie es zum Abendbrot mit nach Hause nehmen würden und wie gern sie das Kind bei sich übernachten lassen würden (Lewis & Lewis, 1987).

Die Teilnehmer antworteten auf einer fünfstufigen Likertskala, die mit Smilies illustriert wurde, die anzeigen, wie gern der Proband das jeweilige Verhalten gegenüber dem Zielkind zeigen würde. Die Skala reicht von 1 = „überhaupt nicht“ (sehr trauriger Smiley) bis 5 = „sehr gern“ (sehr fröhlicher Smiley). Zum dritten Messzeitpunkt wurde zusätzlich zu den beiden bekannten Szenarien noch ein weiteres Szenario mit einem russischen Kind eingefügt.

Auswertung. Für die weiteren Auswertungen wurden Mittelwerte über die einzelnen Items zu jedem Zielkind berechnet. Dabei wurde allerdings das Item „Wie sehr würdest Du ... mögen?“ nicht mit einbezogen, da dieses Item gesondert im

Bereich gruppenbezogene soziale Einstellungen mit affektivem Schwerpunkt ausgewertet wurde. Um Antworttendenzen aus den Mittelwerten herauszurechnen, wurden die Verhaltensintentionen gegenüber den Zielkindern der Fremdgruppen jeweils zu den Intentionen gegenüber dem deutschen Zielkind in Beziehung gesetzt – also eine Differenz zwischen den Werten für das deutsche Kind und den Werten für die Zielkinder der Fremdgruppe berechnet. Somit ergeben sich auch hier Biaswerte für das asiatische Kind und zu t3 auch für das russische Kind. Positive Werte indizieren hier einen stärkeren pro-deutschen Bias. Der Wertebereich der IBM-Biasvariable liegt zwischen -4 (pro Fremdgruppen-Zielkind) und +4 (pro Eigengruppen-Zielkind).

Methodische Qualität. Cameron und Kollegen (2006) geben die Reliabilitäten für die beiden Skalen mit $\alpha = .89$ für das englische Zielkind und mit $\alpha = .90$ für das Flüchtlingskind an. In der vorliegenden Studie liegen die internen Konsistenzen der Skalen ähnlich hoch zwischen $\alpha = .87$ bis $\alpha = .91$ sowohl für das deutsche als auch für das asiatische oder russische Flüchtlingskind.

III.6.7. Interpersonale soziale Einstellungen und Sozialverhalten

Toleranz – Fragebogen zur interpersonalen Toleranz

Ursprung. Der Fragebogen zur interpersonalen Toleranz wurde im Rahmen dieses Projekts 2006 entwickelt. In der ursprünglichen Fassung hatte der Fragebogen 26 Items, die tolerante Einstellungen aus den Bereichen kognitive Aspekte, Verhaltensbasis und Verhaltensintention erfassen sollten. Nach einer Pilotstudie an einer für die vorliegende Studie altersgerechten Jenaer Stichprobe wurden jedoch nur zwei Faktoren extrahiert und die Itemanzahl auf 17 gekürzt.

Inhalt. Die beiden Faktoren wurden als „Akzeptanz und Zivilcourage“ und „Intoleranz“ bezeichnet. Faktor 1 Akzeptanz beinhaltet neun Items, die inhaltlich die Bereiche Meinungs- und Verhaltensfreiheit und Solidarität/Zivilcourage abdecken. Beispielitems für den Faktor Akzeptanz lauten: „Wenn jemand anders denkt als ich, ist das völlig O.K.“ oder „Ich würde auf jeden Fall zu jemandem stehen, der wegen seines Aussehens oder seines Verhaltens ausgelacht wird.“ Auf Faktor 2 Intoleranz laden acht Items hoch, die inhaltlich eine gewisse Rigidität in der eigenen Meinung, Anpassungsbereitschaft an die Gesellschaft und Schwarz-Weiß-Denken abbilden. Beispielitems für den Faktor Intoleranz lauten: „Wenn jemand etwas anderes denkt

als ich, kann ich das nicht annehmen.“ Oder „Wenn sich jemand anders verhält, als die meisten Menschen, finde ich das nicht gut.“

Durchführung. Die Items der beiden Faktoren wurden mit den zehn Items der Skala für interpersonale Ambiguitätsintoleranz gemischt im Fragebogen der Gruppenbefragung dargeboten. Dabei hatten die Kinder fünf Antwortmöglichkeiten die von 1 = „stimmt nicht“ bis 5 = „stimmt genau“ reichten. Zur Illustration wurden die Antwortmöglichkeiten mit Luftballons markiert, die mit dem Grad der Zustimmung größer wurden. Die Kinder erhielten zuvor die Instruktion, zu jedem Item anzukreuzen, wie sehr dieser Satz für sie stimmt. In die weiteren Analysen gehen für jede Subskala Mittelwerte der Skalen mit einem Wertebereich zwischen 1 und 5 ein.

Methodische Qualität. Die Reliabilität der Skalen zur interpersonalen Toleranz zum ersten Messzeitpunkt liegt bei der Subskala Intoleranz bei $\alpha = .70$, während die Subskala Akzeptanz nur eine Reliabilität von $\alpha = .49$ erreicht. Die geringe Reliabilität dieser Subskala ist auf einen deutlichen Deckeneffekt fast aller zugehörigen Items zurückzuführen. Die Mittelwerte der Items liegen zwischen $M = 3.98$ und $M = 4.69$, wobei vier Items sogar einen Mittelwert größer $M = 4.50$ haben (bei einer Skala von eins bis fünf). Auch die Varianzen der Items sind eher gering, so dass die Interkorrelationen zwischen den Items und natürlich die Zuverlässigkeit der Skala niedrig ausfallen. Da durch das Herausnehmen der drei gering ladenden Items die Zuverlässigkeit der Skala auch nicht erhöht wird, verbleiben zur weiteren Auswertung alle Items in der Skala.

Eine ähnliche Verteilung der Reliabilitäten findet sich auch in den Daten des zweiten Messzeitpunktes. Hier liegt die Reliabilität der Subskala Intoleranz bei $\alpha = .73$ und die Reliabilität der Subskala Akzeptanz bei $\alpha = .49$. Hier findet sich ein noch stärkerer Deckeneffekt als zum ersten Messzeitpunkt. Die Reliabilitäten zum dritten Messzeitpunkt liegen bei $\alpha = .56$ für die Subskala Akzeptanz und bei $\alpha = .75$ für die Subskala Intoleranz. Somit ist die Zuverlässigkeit der Intoleranz-Skala als zufriedenstellend einzustufen, die Zuverlässigkeit der Akzeptanz-Skala aber nur noch als ausreichend zu bewerten.

Einschätzungen des Sozialverhaltens durch Eltern und Lehrer

– Strengths and Difficulties Questionnaire

Ursprung: In der vorliegenden Arbeit wurde die deutsche Version (Klasen, Woerner, Rothenberger & Goodman, 2003) des Strengths and Difficulties

Questionnaire (SDQ) von Goodman (1997) sowohl in der Eltern- als auch in der Lehrerversion eingesetzt. Bei diesem Screeningfragebogen geht es darum, dass enge Bezugspersonen das Verhalten des Kindes einschätzen, um Bereiche mit Verhaltensproblemen identifizieren zu können.

Inhalt und Durchführung. Informanten für die Verhaltenseinschätzungen waren hauptsächlich die leiblichen Eltern und die Klassenlehrer der Kinder. Im Falle von entsprechenden Familienverhältnissen wurden auch andere Erziehungsberechtigte (Stiefeltern, Großeltern, etc.) um ihre Einschätzung gebeten.

Dazu wurden den Beurteilern in der Originalfassung 25 Fragen zum Verhalten des Kindes aus fünf verschiedenen Bereichen vorgelegt. Die Bereiche heißen „Emotionale Probleme“, „Verhaltensprobleme“, „Hyperaktivität“, „Verhaltensprobleme mit Gleichaltrigen“ und „prosoziales Verhalten“. Zu jedem dieser Bereiche gibt es fünf Items, bei denen von den Beurteilern eingeschätzt werden soll, in wie weit diese Verhaltensbeschreibung auf das Verhalten des Kindes in den letzten vier Wochen zutrifft oder nicht. Beispielitems aus den einzelnen Bereichen lauten:

„Oft unglücklich oder niedergeschlagen; weint häufig“ (emotionale Probleme)

„Hat oft Wutanfälle; ist aufbrausend“ (Verhaltensprobleme)

„Ständig zappelig“ (Hyperaktivität)

„Wird von anderen gehänselt oder schikaniert“ (Verhaltensprobleme mit Gleichaltrigen)

„Lieb zu jüngeren Kindern“ (prosoziales Verhalten)

Dazu gibt es eine dreigestufte Antwortskala: 0 = „trifft nicht zu“, 1 = „trifft teilweise zu“ und 2 = „trifft vollständig zu“. Da für die vorliegende Arbeit und die Evaluation des Interventionsprogramms der Bereich der Hyperaktivität nicht relevant war, wurde er für diese Studie aus dem SDQ ausgeschlossen. Somit ergibt sich eine Gesamtanzahl von 20 Items aus den vier übrigen Verhaltensbereichen.

Auswertung. Zu jeder dieser vier übrigen Subskalen wurde ein Summenwert berechnet, der die Verhaltenseinschätzung in dem jeweiligen Bereich widerspiegelt. Schließlich wurde ein Gesamtproblemwert berechnet, in den die Summenwerte aus allen Bereichen, bis auf das prosoziale Verhalten, einfließen. Die Summenwerte der Subskalen haben dementsprechend einen Wertebereich von 0 = „keine Probleme“ bis 10 = „sehr problematisches Verhalten“. Der Gesamtproblemwert hat einen Wertebereich von 0 = „keinerlei Probleme“ bis 30 = „äußerst Problematisches Verhalten“. Schließlich bleibt noch der Summenwert des prosozialen Verhaltens, der

nicht in den Gesamtproblemwert mit einfließt und ebenfalls einen Wertebereich von 0 = „kein prosoziales Verhalten“ bis 10 = „höchst prosoziales Verhalten“ aufweist.

Da die Angaben von Vätern und Müttern sehr hoch miteinander korrelieren ($r = .57-.75$) wurden die Werte beider Eltern gemittelt, und werden gemeinsam ausgewertet. Dabei beschränkt sich die Auswertung aus Gründen der Übersichtlichkeit auf den Gesamtproblemwert und die Skala prosozialen Verhaltens.

Methodische Qualität. Die internen Konsistenzen der Skalen und Subskalen erreichen in der vorliegenden Arbeit zufriedenstellende bis gute Werte. Bei Müttern werden Werte von $\alpha = 0.63$ (prosoziales Verhalten t3) bis $\alpha = .80$ (Gesamtproblemwert t3) erreicht. Bei den Vätern liegen die Reliabilitäten etwas darunter und erstrecken sich von $\alpha = .54$ (Verhaltensprobleme t1) bis $\alpha = .78$ (Gesamtproblemwert t2). Bei den Einschätzungen der Lehrer ergeben sich interne Konsistenzen von $\alpha = .63$ (emotionale Probleme t1) bis $\alpha = .83$ (Gesamtproblemwert t2). Somit liegen bei den Beurteilungen der Lehrer die höchsten Reliabilitäten vor. Um den Aufwand für die beteiligten Lehrkräfte zu begrenzen, wurden diese nur zu den ersten beiden Messzeitpunkten um Einschätzungen ihrer Schüler gebeten.

In den unten aufgeführten Tabellen 15 und 16 lassen sich die internen Konsistenzen der eingesetzten Instrumente ablesen.

Tabelle 15

Interne Konsistenzen der Skalen der Gruppen- und Einzelbefragung der Kinder

Skala / MZP	T1	T2	T3
Raven Matrizentest	.43	-	-
Wissenstest (kulturell)	.56	.64	.71
Klassifikation (Personen)	.63	.75	.60
Perspektivenübernahme	.28	.45	.45
Soziale Informationsverarbeitung	.43	.41	.42
MRA (deutsch/positiv)	.82	.84	.80
MRA (russisch/positiv)	.77	.80	.82
MRA (asiatisch/positiv)	.78	.84	.82
MRA (deutsch/negativ)	.83	.83	.82
MRA (russisch/negativ)	.89	.89	.89
MRA (asiatisch/negativ)	.86	.85	.88
Verkuyten (asiatisch)	.56	.60	.64
Verkuyten (deutsch)	.60	.60	.62
Verkuyten (russisch)	-	-	.73
IBM (Lee)	.87	.88	.90
IBM (Renee)	.87	.89	.87
IBM (Vladimir)	-	-	.91
Akzeptanz	.49	.49	.56
Intoleranz	.70	.72	.75

Anmerkungen. Interne Konsistenzen ermittelt über Cronbachs α

Tabelle 16

Interne Konsistenzen der SDQ-Skalen getrennt nach Informanten zu t1, t2 und t3

Informant	SDQ-Skala / MZP	T1	T2	T3
Mutter	Prosoziales Verhalten	.66	.66	.63
	Emotionale Probleme	.72	.64	.68
	Verhaltensprobleme	.61	.66	.63
	Verhaltensprobleme mit Gleichaltrigen	.63	.66	.67
	Gesamtproblemwert	.78	.78	.80
Vater	Prosoziales Verhalten	.63	.66	.63
	Emotionale Probleme	.71	.65	.70
	Verhaltensprobleme	.54	.62	.63
	Verhaltensprobleme mit Gleichaltrigen	.60	.62	.62
	Gesamtproblemwert	.77	.78	.77
Lehrer	Prosoziales Verhalten	.75	.83	-
	Emotionale Probleme	.63	.75	-
	Verhaltensprobleme	.72	.80	-
	Verhaltensprobleme mit Gleichaltrigen	.69	.75	-
	Gesamtproblemwert	.76	.83	-

Anmerkungen. Interne Konsistenzen ermittelt über Cronbachs α

III.7. Operationalisierung der Ziele von PARTS

Aus den theoretischen Grundlagen und empirischen Studien zur Entwicklung negativer gruppenbezogener sozialer Einstellungen wurden in der vorliegenden Arbeit bereits einige Zielkriterien für ein erfolgreiches Interventionsprogramm zur Verbesserung gruppenbezogener sozialer Einstellungen und zur Förderung des allgemeinen Sozialverhaltens und der interpersonellen Toleranz abgeleitet (s. Kapitel II.4. Schlussfolgerungen für PARTS). Im Folgenden Abschnitt wird die Operationalisierung dieser Zielkriterien unter Verwendung der vorliegenden Messinstrumente erläutert.

Theoretische Fundierung

Um sicherzustellen, dass die Interventionsplanung anhand des aktuellen Forschungsstandes erfolgte, muss ein Abgleich der verwendeten Interventionsstrategien mit Schlussfolgerungen aus der Literatur (s. Kapitel II.4. Schlussfolgerungen für PARTS) durchgeführt werden. Eine theoretische Fundierung gilt dann als gewährleistet, wenn die Interventionsmethoden, die Durchführung der Intervention sowie deren begleitende Evaluation dem aktuellen Stand der theoretischen wie empirischen Forschung entspricht.

Implementation

Die Umsetzbarkeit des PARTS-Programms wird in einer separaten Analyse vor der eigentlichen Wirksamkeitsevaluation überprüft. Dafür werden die oben beschriebenen Protokollbögen zur Verhaltenseinschätzung herangezogen. Eine gute Umsetzbarkeit des Programms gilt dann als gewährleistet, wenn die verwendeten Implementationsparameter On-Task-Verhalten und Spaß durchgängig hohe Werte anzeigen, bzw. Off-Task-Verhalten durchgängig niedrige Werte anzeigt.

Proximale Wirksamkeit

Die sozial-kognitiven Fähigkeiten multiple Klassifikation, Perspektivenübernahme und soziales Problemlösen wurden mit Hilfe des Multiple Classification Skill Tasks (Bigler & Liben, 1993), der Chandler Cartoons (Chandler, 1973) und des Tests zum sozialen Problemlösen (Eigenentwicklung, 2007)

operationalisiert. Das interkulturelle Wissen wurde über den Wissenstest (Eigenentwicklung, 2007) gemessen.

Die Herstellung von Stellvertretendem-Kontakt zu russischen Kindern sollte durch die Teilnahme am Programm, speziell durch die Bearbeitung der erfolgreichen Stellvertretender-Kontakt-Geschichten mit den Kindern, sichergestellt sein. Daher erfolgte hier keine separate Messung.

Die proximale Wirksamkeit des Programms gilt dann als gewährleistet, wenn die Interventionsgruppe im Vergleich zur Kontrollgruppe sowohl statistisch als auch praktisch bedeutsame Entwicklungsvorteile, im Sinne einer Förderung, auf den proximalen Effektmaßen aufzeigen. Das bedeutet, dass sich mindestens kleine positive Netto-Effektstärken und signifikante Gruppenunterschiede zwischen den Untersuchungsgruppen auf den proximalen Effektmaßen zeigen müssen.

Distale Wirksamkeit

Die gruppenbezogenen sozialen Einstellungen werden mit Hilfe von Biasmaßen verschiedener Einstellungsinstrumente gemessen. Mit den verwendeten Instrumenten werden drei unterschiedliche Einstellungskomponenten gegenüber verschiedenen Fremdgruppen abgebildet (s. Kapitel III.6. Messinstrumente): Die kognitive und die affektive Einstellungskomponente sollen durch PARTS gleichermaßen positiv beeinflusst werden. Dazu kommt die konative Komponente, welche durch PARTS auf Grund des geringen Anteils an direkten Verhaltensübungen etwas weniger deutlich beeinflusst werden dürfte.

Da keine Instrumente mit Cut-Off-Kriterien vorliegen und eingesetzt werden, und da es interindividuell unterschiedliche Bewertungsstile und Einstellungsausprägungen gibt, werden die gruppenbezogenen sozialen Einstellungen über Relativwerte operationalisiert. Dazu wird die Einstellung gegenüber der Eigengruppe mit der Einstellung gegenüber der jeweiligen Fremdgruppe in Bezug gesetzt (Allport, 1954; Brewer, 1999). Ein solches Vorgehen ist auch deshalb sinnvoll, weil direkte negative Äußerungen gegenüber anderen in dieser Altersgruppe relativ selten sind (Nesdale, 2004). Die beschriebenen Relativwerte werden auch als Biaswerte oder Einstellungsverzerrungen bezeichnet (s.o. Gruppenbezogene soziale Einstellungen).

Eine positive Veränderung der gruppenbezogenen sozialen Einstellungen durch PARTS gilt dann als gewährleistet, wenn die Interventionsgruppe im Vergleich

zur Kontrollgruppe sowohl statistisch als auch praktisch bedeutsame Entwicklungsvorteile, im Sinne geringerer bzw. abnehmender Biaswerte auf den entsprechenden distalen Effektmaßen aufzeigen.

Das allgemeine Sozialverhalten und die interpersonale Toleranz der Kinder wurden anhand der Toleranzskala (Eigenentwicklung, 2007) und den Strengths and Difficulties Questionnaire (SDQ, Goodman, 1997) erfasst.

Eine Verbesserung des allgemeinen Sozialverhaltens und der interpersonalen Toleranz durch PARTS gilt dann als gewährleistet, wenn die Interventionsgruppe im Vergleich zur Kontrollgruppe sowohl statistisch als auch praktisch bedeutsame Entwicklungsvorteile, im Sinne einer Förderung von Akzeptanz und prosozialem Verhalten sowie einer Hemmung von Intoleranz und problematischem Sozialverhalten, auf den entsprechenden distalen Effektmaßen aufzeigen.

Generalisierbarkeit des Programms

Die Generalisierbarkeit des Programms wird überprüft, indem die gruppenbezogenen sozialen Einstellungen in Bezug auf unterschiedliche Gruppen erhoben und die Auswirkungen von PARTS auf die Einstellungen gegenüber diesen Gruppen verglichen werden. Die Generalisierbarkeit des Programms gilt dann als gewährleistet, wenn die Interventionsgruppe im Vergleich zur Kontrollgruppe sowohl statistisch als auch praktisch bedeutsame Entwicklungsvorteile gegenüber der russischen Zielfremdgruppe als auch gegenüber der chinesischen Fremdgruppe aufzeigen.

Allgemeine Wirksamkeit

Die Unabhängigkeit der Wirksamkeit des Programms von differentiellen Einflussfaktoren wird mit Hilfe von Kovarianzanalysen getestet. Eine solche Unabhängigkeit gilt dann als gewährleistet, wenn unter Hinzunahme der entsprechenden Kovariaten keine bedeutsamen Veränderungen der generell gefundenen Interventionseffekte auftreten.

Zeitliche Perspektive

Um die zeitliche Dynamik der Auswirkungen von PARTS zu untersuchen werden die Effekte des Programms in einem Prä-Post-Vergleich und in einem Prä-

Follow-up-Vergleich getestet. Mittels post-hoc-Kontrasten zwischen den Messzeitpunkten wird dabei der Zeitraum der statistisch bedeutsamen Unterschiede identifiziert.

III.8. Hypothesen der PARTS-Studie

Aus den Zielsetzungen und den Operationalisierungen der Konstrukte lassen sich eine Reihe von Hypothesen ableiten. In den folgenden zwei Tabellen werden zunächst die generellen Hypothesen zu den proximalen, dann zu den distalen Zielen dargestellt. Dazu wird die erwartete Richtung der Netto-Effektstärken, also der Differenzen der Prä-Post- oder Prä-Follow-up-Differenzen von Interventions- und Kontrollgruppe aufgelistet. Positive Netto-Effektstärken zeigen dabei eine Steigerung der Werte der Interventionsgruppe im Vergleich zur Kontrollgruppe an. Negative Netto-Effektstärken zeigen dagegen eine Verminderung der Werte der Interventionsgruppe im Vergleich zur Kontrollgruppe an. Je nach abhängigem Effektmaß ist entweder eine positive oder eine negative Netto-Effektstärke ein erwünschter Programmeffekt.

Tabelle 17

Hypothesen der PARTS-Studie zu proximalen Effektmaßen

Konstrukt	Skala	Netto-Effektsärke
Interkulturelles Wissen	Wissenstest	+
Multiple Klassifikationsfähigkeit	Klassifikationstest	+
Perspektivenübernahmefähigkeit	Chandler Cartoons	+
Soziale Informationsverarbeitung	Test zum sozialen Problemlösen	+

Anmerkungen. Netto-Effektstärke = Vergleich der Mittelwertdifferenzen prä-post oder prä-follow-up von Interventions- und Kontrollgruppe; + = positiver Effekt in der Interventionsgruppe, verglichen mit der Kontrollgruppe.

Tabelle 18

Hypothesen der PARTS-Studie zu distalen Effektmaßen

Konstrukt	Skala	Netto-Effektstärke
Gruppenbezogene soziale Einstellungen (kognitiv)	MRA (Person)	-
	Verkuyten-Skala (Gruppe)	-
Gruppenbezogene soziale Einstellungen (affektiv)	Sympathie (Person)	-
	Sympathie (Gruppe)	-
	Haus-Aufgabe (soziale Distanz)	-
Gruppenbezogene soziale Einstellungen (konativ)	IBM (Person)	-
Interpersonale Toleranz	Toleranzskala (Akzeptanz)	+
	Toleranzskala (Intoleranz)	-
Sozialverhalten	SDQ prosoziales Verhalten	+
	SDQ Gesamtproblemwert	-

Anmerkungen. Netto-Effektstärke = Vergleich der Mittelwertdifferenzen prä-post oder prä-follow-up von Interventions- und Kontrollgruppe; + = Steigerung in der Interventionsgruppe, verglichen mit der Kontrollgruppe; - = Senkung in der Interventionsgruppe, verglichen mit der Kontrollgruppe. Gruppenbezogene soziale Einstellungen werden als Biasmaße verwendet.

IV. Ergebnisse

Aufbau des Kapitels

Auf den folgenden Seiten werden die Ergebnisse der PARTS-Studie dargestellt. Dazu werden an dieser Stelle zunächst die Auswertungsstrategien der Studie erläutert. Weiterhin wird auf die verwendeten statistischen Methoden und deren Voraussetzungen, sowie den Umgang mit Voraussetzungsverletzungen eingegangen. Dann folgt die Präsentation der Ergebnisse.

Diese besteht aus den zwei großen Bereichen der *Implementationsanalyse* und der *Wirksamkeitsanalyse*. In der Implementationsanalyse wird die Kernfrage beleuchtet wie gut PARTS implementiert, also umgesetzt werden konnte, und wie es von den Kindern aufgenommen wurde. Im Anschluss daran werden die Ergebnisse der Wirksamkeitsanalyse vorgestellt, in der die Wirksamkeit des Programms auf den verschiedenen proximalen und distalen Effektmaßen überprüft wird.

IV.1. Allgemeine Auswertungsstrategien

Implementationsanalyse

Bei der Implementationsanalyse werden zunächst deskriptiv die drei Implementationsparameter On-Task-Verhalten, Off-Task-Verhalten und Spaß in den Sitzungen über alle Interventionssitzungen hinweg dargestellt. Dies wird ebenso wie alle anderen Auswertungen zur Implementation getrennt für die beiden Informanten, Lehrer und Trainer, berichtet. Weiterhin werden praktisch und statistisch bedeutsame Unterschiede bei der Implementation einzelner Interventionsbereiche untersucht. Dazu werden deskriptive Statistiken, Effektstärken und Konfidenzintervalle berechnet. Schließlich wird der mögliche Einfluss der erhobenen Kovariaten auf die Implementationsparameter deskriptiv (Mittelwerte, Effektstärken und Korrelationen) und inferenzstatistisch (multivariate und univariate Varianzanalysen) geprüft. Somit ergibt sich ein differenziertes Gesamtbild der Implementation.

Wirksamkeitsanalyse

Die Auswertung findet getrennt nach Variablen statt. Dabei werden zunächst die proximalen Skalen und dann die distalen Effektmaße betrachtet.

Zunächst werden dazu die deskriptive Statistik jeder Variable, aufgeschlüsselt nach Interventions- und Kontrollgruppe, sowie die zugehörigen Nettoeffektstärken berichtet. Im Anschluss daran finden sich die inferenzstatistischen Ergebnisse der Varianzanalysen mit und ohne Messwiederholung zur Prüfung auf statistische Signifikanz der vorgefundenen Gruppenunterschiede.

Die deskriptiven und die inferenzstatistischen Analysen wurden zunächst für die gesamte Stichprobe und dann für die jeweilige „Risikogruppe 25“ berechnet. Die Risikogruppe 25 sind die Teilnehmer, die in einer Skala zum Prätest besonders „schlechte“ Werte erzielten und im untersten respektive obersten Viertel der Verteilungen der Skala liegen. Das erste, beziehungsweise dritte, Quartil dient dabei als Trennwert. Da jedoch immer eine unterschiedlich große Anzahl von Teilnehmern auf dem Wert des jeweiligen Quartils liegt, variiert die Teilnehmerzahl von Risikogruppe zu Risikogruppe leicht. Die Risikogruppen 25 wurden explorativ bei den Auswertungen mit betrachtet, um die Wirksamkeit des Programms bei Kindern mit Defiziten in dem entsprechenden Bereich beurteilen zu können. Da jedoch nicht alle Instrumente zum ersten Messzeitpunkt verwendet wurden, konnten nicht in allen Bereichen Risikogruppen bestimmt werden. Dementsprechend wurden nur Analysen zu Skalen, die zu allen drei Messzeitpunkten eingesetzt wurden, durchgeführt. Grundsätzlich zeigen sich in den Risikogruppen 25 ähnliche Ergebnismuster wie in den globalen Auswertungen. Daher finden sich die Ergebnisse dieser Analysen gesammelt in Anhang D.

Es wird bei allen Analysen, sowohl zur Wirksamkeit als auch zur Implementation von PARTS, die Auswirkung möglicher Kovariaten betrachtet. Hierbei handelt es sich hauptsächlich um einige sozio-demographische Variablen, wie das Geschlecht des teilnehmenden Kindes, das Alter des Kindes oder den Entwicklungskontext „Stadt versus Land“; aber auch die zum Prätest gemessene Intelligenz des Kindes geht als Kovariate in die Analysen ein. Da keine spezifischen Hypothesen zur Auswirkung der möglichen Einflussfaktoren vorliegen, werden Ergebnisse bezogen auf die Kovariaten nur dann berichtet, wenn durch den Einfluss der einzelnen Kovariaten bedeutsame Unterschiede zu den globalen

Evaluationsergebnissen, also möglichen statistisch bedeutsamen Interaktionen der Faktoren Messzeitpunkte und Gruppe, bestehen.

Diese Unterschiede werden berichtet, um eine möglichst differenzierte Evaluation von PARTS zu ermöglichen. Auf das Berichten der übrigen Kovarianzanalysen wird aus Gründen des Umfangs und der Übersichtlichkeit verzichtet.

Zum Umgang mit Prätest-Unterschieden der abhängigen Variablen

Da die Zuteilung der Teilnehmer zu den beiden Untersuchungsgruppen bedingt randomisiert erfolgte, ist die Wahrscheinlichkeit für bedeutsame Prätest-Unterschiede auf relevanten Maßen sehr gering. Allerdings sind solche Unterschiede auch nicht ganz ausgeschlossen. Daher wird bei allen Auswertungen der Wirksamkeitsanalyse mittels ANOVA auch auf statistisch bedeutsame Prätest-Unterschiede getestet. Im Falle von signifikanten Prätest-Unterschieden auf zu prüfenden Effektmaßen werden diese Unterschiede gesondert durch eine Moderationsanalyse via lineare Regressionen auf ihren Einfluss auf den Posttest und das Follow-up hin getestet und bei der Interpretation der Ergebnisse berücksichtigt (Baron & Kenny, 1987).

IV.2. Statistische Methoden

IV.2.1. Deskriptive Statistik

Zur Analyse der Daten der PARTS-Studie werden sowohl deskriptive als auch inferenzstatistische Auswertungsverfahren genutzt. Für die deskriptive Analyse der Daten werden die Mittelwerte und Standardabweichungen sowie Korrelationen der Skalen und Items zu den drei Messzeitpunkten berichtet. Diese werden bei der Wirksamkeitsanalyse getrennt nach Interventions- und Kontrollgruppe tabellarisch aufgeführt. Außerdem werden Netto-Effektstärken für die Beurteilung der praktischen Bedeutsamkeit von Mittelwertsunterschieden, wie kurzfristigen und langfristigen Gruppenunterschieden auf den proximalen und distalen Effektmaßen berechnet. Dazu werden folgende Formeln zur Berechnung der Netto-Effektstärken verwendet (Bortz & Döring, 2006):

Netto-Effektstärke im Prä-Post-Vergleich:

$$d_{t2-t1} = \frac{(M_{IGt2} - M_{IGt1}) - (M_{KGt2} - M_{KGt1})}{\frac{(S_{IGt1} + S_{KGt1})}{2}}$$

Netto-Effektstärke im Prä-Follow-Up-Vergleich:

$$d_{t3-t1} = \frac{(M_{IGt3} - M_{IGt1}) - (M_{KGt3} - M_{KGt1})}{\frac{(S_{IGt1} + S_{KGt1})}{2}}$$

Legende:

M_{IG} = Mittelwert Interventionsgruppe

M_{KG} = Mittelwert Kontrollgruppe

S_{IG} = Standardabweichung Interventionsgruppe

S_{KG} = Standardabweichung Kontrollgruppe

Mittels der Netto-Effektstärke kann die Entwicklung der Interventionsgruppe auf einer Variablen oder Skala im Verhältnis zur Entwicklung der Kontrollgruppe auf dieser Skala durch einen Kennwert ausgedrückt werden. Diese Netto-Effektstärke kann zur Quantifizierung der Veränderung zwischen Prä- und Posttest, sowie zwischen Prätest und Follow-up genutzt werden. Für die Skalen und Variablen, die nicht zu allen Messzeitpunkten vorliegen, werden Effektstärken innerhalb eines Messzeitpunktes berechnet. Dazu wird die folgende Formel verwendet:

Effektstärke innerhalb eines Messzeitpunktes (Legende s.o.):

$$d_{t1} = \frac{(M_{IGt1} - M_{KGt1})}{\frac{(S_{IGt1} + S_{KGt1})}{2}}$$

Die Effektstärken ermöglichen es, die praktische Bedeutsamkeit der Interventionseffekte einschätzen zu können. Dabei kann man die Größe der Effektstärken nach Cohen (1992) in kleine Effekte ab $d = 0.2$, mittlere Effekte ab $d = 0.5$ und große Effekte ab $d = 0.8$ einteilen.

IV.2.2. Inferenzstatistik

Die inferenzstatistische Analyse dient zur Überprüfung der vorliegenden Daten auf statistisch bedeutsame Mittelwertsunterschiede zwischen den Untersuchungsgruppen bzw. auf bedeutsame unterschiedliche Veränderungen zwischen den Untersuchungsgruppen. Hierzu werden überwiegend Varianzanalysen mit Messwiederholung berechnet, da es sich bei der PARTS-Studie um eine längsschnittliche Untersuchung der immer gleichen Stichprobe handelt. Post-hoc werden bei signifikanten Interaktionen zusätzlich Kontraste berechnet, um Hinweise zu erhalten, zwischen welchen Messzeitpunkten bedeutsame Unterschiede vorliegen. Als bedeutsamer Indikator für einen Interventionseffekt entsprechend der PARTS-Hypothesen gelten Interaktionen zwischen den unabhängigen Variablen der Gruppenzugehörigkeit und des Messzeitpunktes in die erwartete Richtung. Bei der Auswertung werden bei den Varianzanalysen mit Messwiederholungen zunächst die multivariaten Tests herangezogen, da diese weniger Voraussetzungen benötigen.

Bei Skalen, die nicht zu allen Messzeitpunkten vorliegen, werden Varianzanalysen für unabhängige Stichproben innerhalb eines Messzeitpunktes

berechnet (ANOVAs oder MANOVAs), um die statistische Bedeutsamkeit von gefundenen Mittelwertsunterschieden beurteilen zu können.

IV.2.3. Voraussetzungen der angewendeten inferenzstatistischen

Verfahren

Die Analyse der Daten zur Prüfung der Hypothesen der vorliegenden Arbeit wurden, ebenso wie die Untersuchung der Faktorenstruktur und die Prüfung der internen Konsistenz der eingesetzten Skalen, mittels des Programms PASW Statistics (SPSS 17) durchgeführt. Die angewandten statistischen Auswertungsverfahren beruhen auf dem Allgemeinen Linearen Modell. Vorrangig wurden Varianzanalysen mit Messwiederholung angewendet. Eine Ausnahme bildeten Fälle, in denen die abhängigen Variablen nicht zu allen Messzeitpunkten vorlagen. In diesen Fällen kamen einfache oder multivariate Varianzanalysen für unabhängige Stichproben innerhalb der einzelnen Messzeitpunkte zum Einsatz. Diese Analysemethoden wurden ausgewählt, da sie sich besonders dazu eignen, Mittelwertsunterschiede zwischen Gruppen auf ihre Bedeutsamkeit hin zu prüfen und dabei die Alpha-Fehler-Kumulierung zu minimieren. Beide inferenzstatistischen Verfahren haben ähnliche Voraussetzungen zu ihrer sinnvollen Anwendung (Bortz, 1999; Nachtigall & Wirtz, 2002).

Normalverteilung innerhalb der Gruppen

Diese Voraussetzung beinhaltet, dass die Fehlerkomponente, also die Abweichung eines Messwertes vom jeweiligen Stichprobenmittel, innerhalb jeder Treatmentstufe (z.B. Experimental- vs. Kontrollgruppe), beziehungsweise jeder Faktorstufenkombination bei mehrfaktoriellen Analysen, normalverteilt ist. Bei Verletzungen dieser Voraussetzung reagiert der F-Test, der signifikante Abweichungen zwischen den Gruppenmittelwerten testet, je nach Form der Fehlerverteilung unterschiedlich: Bei schiefen Verteilungen ist der Test robust, bei sehr schmalgipfligen Verteilungen bleibt der Test eher konservativ, und bei extrem breitgipfligen Verteilungen reagiert er progressiver.

Varianzhomogenität

Die zweite Voraussetzung der Varianzanalyse fordert gleiche Varianzen der Fehler in den verschiedenen Gruppen. Wenn die Varianzen ungleich sind, aber die Gruppengrößen etwa gleich sind, bleibt der F-Test auch hier robust. Wenn allerdings die Gruppengrößen ungleich sind und die Varianzhomogenität verletzt ist, besteht die Gefahr, dass der Test progressiver reagiert, also eher ein signifikantes Ergebnis anzeigt.

Zirkularitätsannahme

Bei der Varianzanalyse mit Messwiederholung müssen als weitere Voraussetzung die Korrelationen zwischen den Faktorstufen des Messwiederholungsfaktors homogen sein. Eine Verletzung dieser Voraussetzung führt zu progressiven Entscheidungen.

Unabhängige Stichproben

Diese Voraussetzung verlangt, dass die Fehlerkomponenten der einzelnen Gruppen unabhängig voneinander sein müssen. Dies wird zum Beispiel durch eine zufällige Zuweisung der Probanden zu Untersuchungsgruppen (Randomisierung) gewährleistet, wobei sichergestellt werden muss, dass jeder Versuchsteilnehmer nur einem Treatment ausgesetzt wird.

IV.2.4. Prüfung der Voraussetzungen und Umgang mit**Voraussetzungsverletzungen****Normalverteilung innerhalb der Gruppen**

Die Überprüfung der Normalverteilungsvoraussetzung wird, auf Grund der Robustheit des F-Tests gegenüber Verletzungen in diesem Bereich, nicht weiter verfolgt. Zudem kann nach dem zentralen Grenzwertsatz davon ausgegangen werden, dass ab einer Gruppengröße von $n > 30$ Probanden die erfassten Mittelwerte normalverteilt sind (Nachtigall, 2002).

Varianzhomogenität

Die Voraussetzung der Varianzhomogenität wird bei jeder Varianzanalyse mittels des Levene-Tests auf Gleichheit der Fehlervarianzen mitgeprüft. Bei den

multivariaten Varianzanalysen erfolgt zusätzlich der Box-M-Test auf Gleichheit der Kovarianzmatrizen zwischen den Gruppen. Da bei diesen Tests die Null-Hypothese die Wunschhypothese ist, wird das Signifikanzniveau hier auf $\alpha = 0.25$ hochgesetzt. Falls der Levene-Test oder der Box-M-Test signifikant wird, die Fehlervarianzen beziehungsweise Kovarianzen zwischen den Gruppen also nicht gleich sind, wird dies nur berichtet, falls außerdem die Gruppengrößen unterschiedlich sind. Allein in diesem Fall wird der F-Test unzuverlässig und reagiert zu progressiv, so dass das entsprechende Testergebnis mit Vorsicht interpretiert werden muss. In den übrigen Fällen werden die Ergebnisse des Levene-Tests und des Box-M-Tests nicht berichtet.

Zirkularitätsannahme

Verletzungen der Zirkularitätsannahme liegen vor, wenn die Korrelationen zwischen den Messzeitpunkten unsystematisch variieren und sich disproportional zur Einheitsmatrix verhalten. Dies wird mit dem Mauchly-Test auf Sphärizität für jede Varianzanalyse mit Messwiederholungen geprüft. Da auch hier die Null-Hypothese die Wunschhypothese ist, wird ebenfalls das Signifikanzniveau auf ein $\alpha = 0.25$ hochgesetzt. Die sich aus der Verletzung der Zirkularitätsannahme ergebenden progressiven Entscheidungen lassen sich durch eine Modifizierung der Freiheitsgrade des F-Tests korrigieren. In diesem Fall wird die korrigierte Prüfgröße von Greenhouse und Geisser (1958) angewendet, um weiterhin auf einem Alpha-Niveau von 5 % testen zu können. Da die korrigierte Prüfgröße ein robustes Testen ermöglicht, werden die Ergebnisse zur Zirkularitätsannahme, abgesehen von der verwendeten Prüfgröße, nicht weiter berichtet.

Unabhängige Stichproben

Die Voraussetzung der unabhängigen Stichproben beziehungsweise der unabhängigen Fehlerkomponenten ist durch die klassenweise zufällige Zuweisung der Probanden zu den einzelnen Versuchsgruppen gewährleistet.

Zum Umgang mit der Alpha-Fehler-Kumulierung

In der vorliegenden Arbeit wird eine große Anzahl an statistischen Signifikanztests durchgeführt. Dies führt zu einem kumulierten Alpha-Fehler und erhöht die Gefahr rein zufällig entstandene Unterschiede als signifikant zu

interpretieren, bzw. die Irrtumswahrscheinlichkeit hierfür nicht mehr genau angeben zu können. Daher wird üblicherweise das Alpha-Niveau bei der Anwendung vieler Tests neu berechnet und herabgesetzt (beispielsweise mit der Bonferoni-Korrekturformel). Dieses Vorgehen würde in der vorliegenden Studie dazu führen, dass das korrigierte Alpha-Niveau unter einem Werte von $\alpha = 0.001$ liegen würde. Da in der vorliegenden Evaluationsstudie jedoch eher kleine Interventionseffekte erwartet werden, erweist sich diese Vorgehensweise in diesem Fall als nicht sinnvoll. Es könnte dazu führen, dass die kleinen aber bedeutsamen Veränderungen durch das Programm nicht als solche indentifiziert werden (Nachtigall & Wirtz, 2002). Aus diesem Grund wird in der vorliegenden Arbeit auf eine Alpha-Fehler-Korrektur verzichtet.

IV.3. Implementationsanalyse

Die Implementationsanalyse untersucht die Beteiligung der Kinder am Präventionsprogramm. Dazu werden drei Parameter, die von den beteiligten Trainern und den begleitenden Klassenlehrern eingeschätzt wurden, betrachtet: Das On-Task-Verhalten, das Off-Task-Verhalten und der Spaß an der jeweiligen Interventionseinheit. Daraufhin werden die oben genannten Parameter separat für die einzelnen Programmteile Stellvertretender-Kontakt, Wissensintervention und die Förderung sozial-kognitiver Fähigkeiten analysiert. Somit wird die Implementation der einzelnen Programmbestandteile differenziert ausgewertet. Außerdem werden die Kinder identifiziert, die sich dauerhaft nicht adäquat am Programm beteiligten, also überdurchschnittlich hohes Off-Task-Verhalten über mehrere Sitzungen hinweg zeigten. Schließlich wird die Implementation auf Zusammenhänge mit einigen möglichen Kovariaten geprüft.

IV.3.1. Implementation des Gesamtprogramms

Wie aus Tabelle 19 zum *On-Task-Verhalten* der Teilnehmer ersichtlich, liegen die Werte der Kinder hier sehr hoch. Die Mittelwerte für das On-Task-Verhalten liegen in einem Range von 1.79 (Lektion 8 Trainereinschätzung) bis hin zu 2.34 (Lektion 1 und 2 Trainereinschätzung) bei einer Skala von 0 = „gar nicht“ bis 3 = „sehr stark“ (s. Kapitel III.6. Messinstrumente). Im Mittel über alle Sitzungen hinweg liegt die Lehrereinschätzung für das On-Task-Verhalten bei $M = 2.13$ ($SD = 0.51$) und für die Trainereinschätzungen bei $M = 2.00$ ($SD = 0.54$). Die Korrelationen zwischen den Lehrer- und Trainereinschätzungen liegen im kleinen bis mittleren Größenbereich zwischen $r = .27$ (Lektion 5) und $r = .49$ (Lektion 15). Im Schnitt korrelieren die Einschätzungen zu $r = 0.39$.

In Tabelle 20 sind die Mittelwerte zum *Off-Task-Verhalten* der Teilnehmer abgebildet. Sowohl die Trainer- als auch die Lehrereinschätzungen ergeben hier nur sehr geringe Mittelwerte. Der Wertebereich erstreckt sich von $M = 0.18$ (Lektion 14 Lehrereinschätzung) bis hin zu $M = 0.56$ (Lektionen 4 und 9 Lehrereinschätzung) bei einer Skala von 0 = „gar nicht“ bis 3 = „sehr stark“. Im Mittel über alle Sitzungen hinweg liegen die Lehrereinschätzungen des Off-Task-Verhaltens bei $M = 0.46$

($SD = 0.52$) und die Trainereinschätzungen bei $M = 0.33$ ($SD = 0.39$). Die Korrelationen zwischen Lehrer- und Trainereinschätzungen liegen hier, ähnlich wie für das On-Task-Verhalten, im kleinen bis mittleren Bereich. Sie reichen von $r = .21$ (Lektion 3) bis zu $r = .54$ (Lektion 9). Im Durchschnitt korrelieren die Einschätzungen von Lehrern und Trainern zu $r = .37$.

Tabelle 21 schließlich führt die Mittelwerte der Lehrer- und Trainereinschätzungen zum *Spaß* der Kinder an der jeweiligen Lektion auf. Hier reicht die Skala allerdings nur von 0 = „sehr wenig“ bis 2 = „viel“ (s. Kapitel III.6. Messinstrumente). Die mittleren Einschätzungen für den Spaß der Kinder am Programm sind sehr hoch. Sie reichen von $M = 1.26$ (Lektion 11 Trainer) bis zu $M = 1.73$ (Lektion 3 Lehrer). Die Lehrereinschätzungen haben einen Mittelwert von $M = 1.56$ ($SD = 0.33$) und die Trainereinschätzungen einen Mittelwert von $M = 1.49$ ($SD = 0.31$) über die Zeit. Die Korrelationen zwischen den Einschätzungen von Lehrern und Trainern sind hier allerdings geringer als bei On-Task- und Off-Task-Verhalten. Sie liegen zwischen $r = .14$ und $r = .49$. Im Durchschnitt über alle Sitzungen hinweg liegt die Korrelation bei $r = .29$.

Zur Identifikation von besonders eingängigen oder besonders schwierigen Interventionsbereichen des multimodalen Programms PARTS, werden die einzelnen Interventionsbereiche (Stellvertretender-Kontakt, wissensbasierte Intervention und Förderung sozial-kognitiver Fähigkeiten) im Folgenden getrennt voneinander auf ihre Implementation hin untersucht.

Tabelle 19

On-Task-Verhalten der Teilnehmer über die 15 Lektionen, getrennt nach Lehrer- und Trainereinschätzungen, sowie Korrelationen der Einschätzungen von Lehrern und Trainern

	1.u2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	gesamt
	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)
On-Task- Verhalten Lehrer	2.11 (0.76)	2.33 (0.66)	2.26 (0.74)	2.12 (0.69)	2.01 (0.74)	1.98 (0.74)	2.08 (0.75)	2.07 (0.77)	2.21 (0.81)	2.08 (0.77)	2.20 (0.78)	2.11 (0.80)	2.24 (0.74)	2.10 (0.77)	2.13 (0.51)
On-Task- Verhalten Trainer	2.34 (0.68)	2.19 (0.69)	2.24 (0.77)	2.27 (0.79)	1.90 (0.92)	1.99 (0.80)	1.79 (0.85)	1.99 (0.71)	1.81 (0.81)	1.91 (0.84)	1.81 (0.79)	1.88 (0.85)	1.95 (0.87)	2.09 (0.78)	2.00 (0.54)
Korrelation (Lehrer-Trainer)	.32*	.29*	.33*	.27*	.48*	.35*	.38*	.41*	.40*	.44*	.36*	.48*	.46*	.49*	.73*

Anmerkungen. Wertebereich On-Task-Verhalten: 0 = gar nicht – 3 = sehr stark; * = signifikante Produkt-Moment-Korrelation nach Pearson mit $p < 0.05$.

Tabelle 20

Off-Task-Verhalten der Teilnehmer über die 15 Lektionen, getrennt nach Lehrer- und Trainereinschätzungen, sowie Korrelationen der Einschätzungen von Lehrern und Trainern

	1.u2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	gesamt
	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)
Off-Task-Verhalten Lehrer	0.49 (0.77)	0.47 (0.87)	0.56 (0.74)	0.37 (0.71)	0.48 (0.72)	0.45 (0.65)	0.45 (0.81)	0.56 (0.75)	0.55 (0.81)	0.53 (0.74)	0.39 (0.69)	0.39 (0.59)	0.18 (0.45)	0.45 (0.71)	0.46 (0.52)
Off-Task-Verhalten Trainer	0.38 (0.64)	0.32 (0.63)	0.32 (0.62)	0.28 (0.60)	0.20 (0.54)	0.43 (0.64)	0.31 (0.58)	0.28 (0.60)	0.31 (0.58)	0.37 (0.67)	0.44 (0.70)	0.31 (0.55)	0.42 (0.64)	0.27 (0.56)	0.33 (0.39)
Korrelation (Lehrer-Trainer)	.31*	.21*	.30*	.33*	.24*	.34*	.34*	.54*	.50*	.41*	.48*	.37*	.34*	.41*	.70*

Anmerkungen. Wertebereich On-Task-Verhalten: 0 = gar nicht – 3 = sehr stark; * = signifikante Produkt-Moment-Korrelation nach Pearson mit $p < 0.05$.

Tabelle 21

Spaß der Teilnehmer über die 15 Lektionen, getrennt nach Lehrer- und Trainereinschätzungen, sowie Korrelationen der Einschätzungen von Lehrern und Trainern

	1.u2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	gesamt
	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)
Spaß Lehrer	1.55 (0.54)	1.73 (0.48)	1.65 (0.49)	1.56 (0.53)	1.45 (0.56)	1.47 (0.51)	1.64 (0.48)	1.50 (0.52)	1.67 (0.47)	1.49 (0.50)	1.55 (0.53)	1.57 (0.53)	1.59 (0.52)	1.52 (0.52)	1.56 (0.33)
Spaß Trainer	1.61 (0.51)	1.43 (0.53)	1.63 (0.51)	1.62 (0.49)	1.46 (0.54)	1.46 (0.58)	1.46 (0.55)	1.63 (0.50)	1.55 (0.52)	1.26 (0.53)	1.34 (0.57)	1.31 (0.58)	1.50 (0.57)	1.60 (0.53)	1.49 (0.31)
Korrelation (Lehrer-Trainer)	.21*	.23*	.49*	.37*	.36*	.34*	.14	.25*	.28*	.29*	.23*	.32*	.28*	.28*	.64*

Anmerkungen. Wertebereich Spaß: 0 = gar nicht – 2 = sehr viel; * = signifikante Produkt-Moment-Korrelation nach Pearson mit $p < 0.05$.

Tabelle 22

Mittelwerte, Standardabweichungen, Effektstärken und 95%-Konfidenzintervalle, nach Interventionsbereichen getrennt, im Vergleich zum Gesamtmittelwert der Implementationsparameter

		Wissen	St.-Kontakt	soz.-kog.	gesamt
		<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
		(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)
On-Task-Verhalten	Lehrer	2.17 (0.54) <i>d</i> = 0.08	2.15 (0.58) <i>d</i> = 0.04	2.01* <i>b</i> (0.60) <i>d</i> = -0.22	2.13 (0.51)
	95%KI	2.10-2.24	2.08-2.23	1.93-2.09	2.06-2.20
	Trainer	2.09 (0.59) <i>d</i> = 0.16	1.94 (0.57) <i>d</i> = -0.11	1.92 (0.69) <i>d</i> = -0.13	2.00 (0.54)
	95%KI	2.01-2.17	1.86-2.02	1.83-2.02	1.93-2.07
Off-Task-Verhalten	Lehrer	0.47 (0.61) <i>d</i> = 0.02	0.45 (0.60) <i>d</i> = -0.02	0.47 (0.55) <i>d</i> = 0.02	0.46 (0.52)
	95%KI	0.39-0.53	0.39-0.56	0.37-0.53	0.39-0.53
	Trainer	0.33 (0.44) <i>d</i> = 0.00	0.33 (0.42) <i>d</i> = 0.00	0.33 (0.46) <i>d</i> = 0.00	0.33 (0.39)
	95%KI	0.27-0.39	0.27-0.39	0.27-0.39	0.27-0.38
Spaß	Lehrer	1.60 (0.36) <i>d</i> = 0.12	1.59 (0.37) <i>d</i> = 0.09	1.47* <i>a</i> (0.43) <i>d</i> = -0.23	1.56 (0.33)
	95%KI	1.55-1.65	1.53-1.64	1.41-1.53	1.52-1.61
	Trainer	1.59* <i>a</i> (0.33) <i>d</i> = 0.31	1.44 (0.36) <i>d</i> = -0.15	1.38* <i>b</i> (0.40) <i>d</i> = -0.31	1.49 (0.31)
	95%KI	1.54-1.63	1.39-1.49	1.33-1.44	1.44-1.53

Anmerkungen. Wissen: Lektionen 1&2, 4, 5, 10, 14; Stellvertretender-Kontakt: Lektionen 3, 8, 9, 12, 13. Sozial-kognitiv: Lektionen 6, 7, 11. *d* = Effektstärke im Vergleich zum Zeilenmittel;

*a = signifikanter Unterschied (außerhalb des 95%-Konfidenzintervalls) zu beiden anderen Bereichen.

*b = signifikanter Unterschied (außerhalb des 95%-Konfidenzintervalls) zu einem anderen Bereich.

IV.3.2. Implementation getrennt nach Interventionsbereichen

Um zu prüfen, ob einzelne Interventionsbereiche besonders leicht oder schwierig implementiert werden konnten, wurden in der obigen Tabelle 22 die Implementationsparameter für die Einheiten der einzelnen Bereiche zusammengefasst: Zum einen die Wissensinterventionseinheiten (Einheiten 1 und 2, 4, 5, 10 und 14), zum anderen die Einheiten in denen eine Stellvertretender-Kontakt-Intervention stattfindet (Einheiten 3, 8, 9, 12 und 13) und schließlich die drei Einheiten, die ausschließlich sozial-kognitive Trainingselemente beinhalten (Einheiten 6, 7 und 11). Dazu muss festgehalten werden, dass es sich bei den Einheiten, in denen eine Stellvertretender-Kontakt-Intervention stattfindet, um „gemischte“ Interventionseinheiten handelt, da in diesen Lektionen auch Übungen zur Förderung sozial-kognitiver Fähigkeiten durchgeführt werden (s. Kapitel III.2.3. Umsetzung der Interventionsstrategien). Ein bedeutsamer Teil der Interventionseinheiten wird dort allerdings immer auf den stellvertretenden Kontakt verwendet, so dass sich in der Zusammenfassung dieser Einheiten die Implementation der Stellvertretender-Kontakt-Intervention widerspiegelt. Bei den Wissensseinheiten handelt es sich um „reine Wissensseinheiten“. Eine Ausnahme bildet die Doppeleinheit zu Beginn der Intervention, die außer der Wissensvermittlung noch eine Vorstellung der Trainier und eine Einführung in das Programm PARTS beinhaltet. Hier existiert nur eine zusammenfassende Einschätzung der Implementation für die beiden Lektionen 1 (Einführung) und 2 (1. Wissensinterventionseinheit), da beide Einheiten in einer Doppelstunde direkt hintereinander durchgeführt wurden.

Betrachtet man nun die Implementationsparameter getrennt für die einzelnen Interventionsbereiche, lassen sich einige Unterschiede und Gemeinsamkeiten zwischen den Bereichen feststellen.

Da bei der Prüfung der Mittelwertsunterschiede auf statistische Signifikanz via T-Test für abhängige Stichproben oder Varianzanalyse mit Messwiederholung der Standardfehler der Mittelwertsdifferenzen bei einer Stichprobe von $N = 209$ sehr klein wird, werden hier selbst die geringsten Mittelwertsunterschiede statistisch

bedeutsam. Daher wird an dieser Stelle auf eine inferenzstatistische Prüfung via T-Test oder Varianzanalyse mit Messwiederholung verzichtet. Stattdessen werden die Effektstärken der Einzelmittelwerte im Vergleich zum Gesamtmittelwert der jeweiligen Lehrer- und Trainereinschätzungen berichtet. Außerdem werden die 95%-Konfidenzintervalle der Einzelmittelwerte dargestellt. Bei nicht überschneidenden Konfidenzintervallen der Einzelmittelwerte kann von einem bedeutsamen Mittelwertsunterschied in der Population ausgegangen werden (Nachtigall & Wirtz, 2002).

In Bezug auf das *On-Task-Verhalten* liegen die Interventionen mit ausschließlich sozial-kognitiven Inhalten leicht hinter den anderen beiden Bereichen zurück. In der Lehrereinschätzung wird dieser Unterschied im Vergleich mit dem Wissensinterventionsbereich auch statistisch bedeutsam (s. Konfidenzintervall Tabelle 22). Dieser Mittelwertsunterschied spiegelt sich auch in der negativen Effektstärke im Vergleich zum Gesamtmittelwert der Lehrereinschätzungen des On-Task-Verhaltens wider. In der Trainereinschätzung fällt zudem eine deskriptiv geringere Einschätzung des On-Task-Verhaltens in den Sitzungen mit stellvertretendem Kontakt im Vergleich zur Wissensintervention auf. Diese wird aber nicht statistisch signifikant.

Das *Off-Task-Verhalten* unterscheidet sich zwischen den einzelnen Interventionsbereichen kaum. Hier lässt sich einzig eine im Vergleich zur Trainereinschätzung durchgehend höhere Lehrereinschätzung feststellen. Auch die Konfidenzintervalle überschneiden sich hier fast vollständig, was für etwa gleich große Mittelwerte im Off-Task-Verhalten der einzelnen Interventionsbereiche spricht.

Der *Spaß* der Kinder wurde ähnlich eingeschätzt wie das On-Task-Verhalten. Auch hier zeigt sich ein leichter Abfall der Einschätzungen im sozial-kognitiven Interventionsbereich. Dieser wird in der Lehrereinschätzung auch gegenüber den beiden anderen Interventionsbereichen signifikant (Konfidenzintervalle überschneiden sich nicht). Außerdem schätzten die Trainer den Spaß der Kinder bei Wissensinterventionseinheiten signifikant höher ein, als in den anderen beiden Bereichen (Konfidenzintervalle überschneiden sich nicht).

Natürlich gibt es trotz dieser guten Ergebnisse immer Kinder, die nicht mit einer Intervention wie PARTS erreicht werden können. Diese Kinder können

zumindest teilweise durch ein erhöhtes Off-Task-Verhalten über mehrere Sitzungen hinweg identifiziert werden. Dies wird im Folgenden genauer beschrieben.

IV.3.3. Gruppe der „Störenfriede“

Die Gruppe der Kinder mit besonders hohem Off-Task-Verhalten wurde über die Verteilungen der Off-Task-Variablen ermittelt. Als Auswahlkriterium für hohes Off-Task-Verhalten wurde das 75.-Perzentil der Off-Task-Variablen gewählt. Alle Kinder, die mit ihrem Off-Task-Verhalten über diesem Wert lagen, wurden für diese Lektion zu denjenigen mit hohem Off-Task-Verhalten gezählt. Alle Kinder, deren Werte gleich hoch oder niedriger waren, wurden zu den Kindern mit nicht auffälligem Off-Task-Verhalten gezählt. Da bei jeder der Off-Task-Variablen eine verschieden große Anzahl von Kindern mit ihren Werten auf dem 75.-Perzentil liegt, ergeben sich unterschiedlich große Fallzahlen für die Kinder mit hohem Off-Task-Verhalten zu den einzelnen Sitzungen. Die Anzahl der Kinder mit hohem Off-Task-Verhalten zu jeder Interventionssitzung lässt sich getrennt nach Lehrer- und Trainereinschätzungen in Tabelle 23 ablesen. Die Zahlen schwanken zwischen $n = 9$ (4.3% der an PARTS teilnehmenden Kinder, z.B. in Lektion 13) und $n = 48$ (23.0% der an PARTS teilnehmenden Kinder, in Lektion 8 Trainereinschätzung).

Die Kinder mit hohem Off-Task-Verhalten zu den einzelnen Lektionen sind jedoch nicht immer dieselben. Für eine differenzierte Analyse der Implementation von PARTS ist es wichtig, die Anzahl der Kinder zu identifizieren, die während mehrerer Sitzungen besonders unaufmerksam oder störend aufgefallen sind. Diese Kinder scheinen an dem Programm weniger interessiert oder über- bzw. unterfordert gewesen zu sein. Man kann nicht davon ausgehen, dass die Implementation von PARTS bei diesen Kindern genauso wie bei ihren Klassenkameraden mit unauffälligem Off-Task-Verhalten erfolgreich durchgeführt werden konnte. Daher wurde für jedes Kind die individuelle Anzahl von Sitzungen mit hohem Off-Task-Verhalten ermittelt. Kinder mit einem Wert von sieben oder mehr, die also in mindestens der Hälfte der 14 Interventionssitzungen ein deutlich erhöhtes Off-Task-Verhalten gezeigt haben, wurden als Gruppe der Hoch-Off-Task-Kinder (im Folgenden kurz „Störenfriede“ genannt) zusammengefasst. Wenn man sowohl die Lehrer-, als auch die Trainereinschätzungen bezüglich des hohen Off-Task-Verhaltens betrachtet, ergibt sich eine Gruppe von $n = 26$ Kindern (12.4% der an PARTS teilnehmenden Kinder) mit hohem Off-Task-Verhalten über mindestens sieben Interventionssitzungen hinweg. Davon erfüllen lediglich fünf Kinder (2.4%)

das Auswahlkriterium, sowohl der Lehrer- als auch der Trainereinschätzung zufolge. Die übrigen 21 Kinder sind entweder nur von Lehrern oder nur von Trainern als „Störenfriede“ identifiziert worden.

Von den $n = 26$ „Störenfrieden“ sind $n = 21$ Jungen und $n = 5$ Mädchen. Es sind also trotz einer beinahe ausgeglichenen Geschlechterverteilung in der gesamten Interventionsgruppe viermal so viele Jungen als Mädchen unter den „Störenfrieden“. Die „Störenfriede“ sind zum ersten Messzeitpunkt zwischen 89 und 111 Monaten alt und im Mittel $M = 101.85$ ($SD = 4.93$) Monate alt gewesen. Damit entsprechen die Störenfriede fast exakt dem Altersgesamtdurchschnitt der PARTS-Teilnehmer zum ersten Messzeitpunkt ($M = 101$ Monate $SD = 4.75$).

Tabelle 23

Anzahl der Kinder mit hohem Off-Task-Verhalten, getrennt nach Lehrer und Trainereinschätzungen, über die 15 Lektionen hinweg

	1.u2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
Lehrer	21 (10.0%)	43 (20.6%)	20 (9.6%)	39 (18.7%)	21 (10.0%)	13 (6.2%)	18 (8.6%)	20 (9.6%)	29 (13.9%)	20 (9.6%)	15 (7.2%)	9 (4.3%)	27 (12.9%)	19 (9.1%)
Trainer	13 (6.2%)	46 (22%)	43 (20.6%)	40 (19.1%)	31 (14.8%)	13 (6.2%)	48 (23.0%)	42 (20.1%)	9 (4.3%)	16 (7.7%)	22 (10.5%)	9 (4.3%)	16 (7.7%)	43 (20.6%)

Anmerkungen. *n* = Anzahl der Kinder mit hohem Off-Task-Verhalten in der jeweiligen Lektion; % = prozentualer Anteil der Kinder mit hohem Off-Task-Verhalten an der Gesamtgruppe der an PARTS teilnehmenden Kinder.

IV.3.4. Implementation unter Einbezug der Kovariaten

Im Folgenden werden die Auswirkungen verschiedener möglicher Einflussfaktoren, wie Geschlecht und Entwicklungskontext, und Kovariaten, wie Alter und Intelligenz, auf die Implementation von PARTS geprüft und berichtet.

Geschlecht

Es zeigen sich einige Geschlechtsunterschiede in der Implementation von PARTS. Tabelle 24 zeigt die gemittelten Einschätzungen von Lehrern und Trainern bei Jungen ($n = 103$) und Mädchen ($n = 106$) im Vergleich.

Tabelle 24

Geschlechtsunterschiede der mittleren Implementationsparameter getrennt nach Lehrer- und Trainereinschätzungen

	On-Task		Off-Task		Spaß	
	J	M	J	M	J	M
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Lehrer	2.00	2.24	0.62	0.30	1.50	1.62
	(0.48)	(0.52)	(0.58)	(0.39)	(0.32)	(0.33)
	$d = -0.48^*$		$d = 0.65^*$		$d = -0.37^*$	
Trainer	1.91	2.09	0.43	0.22	1.44	1.53
	(0.53)	(0.54)	(0.46)	(0.27)	(0.32)	(0.30)
	$d = -0.34^*$		$d = 0.56^*$		$d = -0.39^*$	

Anmerkungen. J = Jungen; M = Mädchen; d = Effektstärke im Vergleich Jungen vs. Mädchen; * = signifikanter Unterschied mit $p < 0.05$ zwischen den Mittelwerten im Vergleich Jungen vs. Mädchen.

Die Unterschiede in der Implementation bei Jungen und Mädchen zeichnen sich deutlich sowohl in den Lehrer- als auch in den Trainereinschätzungen ab. Für das On-Task-Verhalten zeigen die Mädchen höhere Werte als die Jungen. Die signifikanten Mittelwertsunterschiede (Lehrer: $F [1, 207] = 12.07$, $p = .001$; Trainer: $F [1, 207] = 6.28$, $p = .010$) haben auch praktische Bedeutsamkeit: sie zeigen eine kleine (Trainereinschätzung) bis mittlere Effektstärke (Lehrereinschätzung).

Für das Off-Task-Verhalten finden sich die größten signifikanten Unterschiede zwischen Jungen und Mädchen (Lehrer: $F [1, 207] = 20.51, p = .000$; Trainer: $F [1, 207] = 16.07, p = .000$). Die Mädchen zeigen ein deutlich geringeres Off-Task-Verhalten als die Jungen.

Beim Spaß schließlich erzielen die Mädchen leicht höhere Werte als die Jungen. Auch dieser kleine Effekt wird statistisch signifikant (Lehrer: $F [1, 207] = 6.62, p = .010$; Trainer: $F [1, 207] = 4.47, p = .040$).

Ähnliche Geschlechtsunterschiede finden sich auch, wenn man die drei Interventionsbereiche (Wissensbasierte Intervention, Stellvertretender-Kontakt und Förderung sozial-kognitiver Fähigkeiten) einzeln betrachtet. Auf eine ausführlichere Darstellung hierzu soll an dieser Stelle verzichtet werden.

Alterseffekte

Auch das Alter der Kinder scheint einen Einfluss auf die Implementation des Programms zu haben. Da Alter eine kontinuierliche Variable ist, wurden zur Prüfung des Einflusses des Alters auf die Implementationsparameter Korrelationen berechnet. Diese sind in Tabelle 25 wiedergegeben.

Tabelle 25

Korrelationen zwischen dem Alter der Kinder zum ersten Messzeitpunkt und den mittleren Implementationsparametern

	On-Task	Off-Task	Spaß
Lehrer	-.14*	.14*	-.16*
Trainer	-.10	.13	-.12

Anmerkungen. * = signifikante Produkt-Moment-Korrelation nach Pearson mit $p < 0.05$

Sowohl in den Lehrer- als auch in den Trainereinschätzungen zeigt sich der gleiche Trend: Je älter die Kinder werden, desto niedriger liegen die Werte beim On-Task-Verhalten und Spaß der Kinder und desto höher liegen die Werte ihres Off-Task-Verhaltens. Dies besagen die negativen Korrelationen zwischen dem Alter der Kinder und dem On-Task-Verhalten sowie dem Spaß, und die positiven Korrelationen zwischen dem Alter und dem Off-Task-Verhalten der Kinder. Allerdings sind die Zusammenhänge sehr klein und auch nur bei den Lehrereinschätzungen

statistisch bedeutsam. Der gleiche Trend zeigt sich auch für die einzelnen Interventionsbereiche. Bei den Lektionen mit Stellvertretender-Kontakt werden die Korrelationen des Alters mit den Lehrereinschätzungen nicht mehr signifikant.

Entwicklungskontext Stadt versus Land

An dieser Stelle wird die Frage geklärt, ob sich das Programm besser bei „Stadtkindern“ als bei „Landkindern“ implementieren lässt. Tabelle 26 gibt die Mittelwerte der Implementationsparameter getrennt für diese beiden Gruppen wider.

Tabelle 26

Mittlere Implementationsparameter für Stadt- versus Landkinder getrennt nach Lehrer- und Trainereinschätzungen

	On-Task		Off-Task		Spaß	
	Stadt	Land	Stadt	Land	Stadt	Land
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Lehrer	2.24	2.09	0.32	0.50	1.52	1.57
	(0.36)	(0.54)	(0.35)	(0.55)	(0.37)	(0.32)
	$d = 0.33$		$d = 0.39^*$		$d = 0.15$	
Trainer	2.13	1.96	0.35	0.32	1.48	1.48
	(0.44)	(0.56)	(0.39)	(0.40)	(0.33)	(0.31)
	$d = 0.34$		$d = 0.08$		$d = 0.00$	

Anmerkungen. d = Effektstärke im Vergleich Stadtkinder vs. Landkinder; * = signifikanter Unterschied mit $p < 0.05$ zwischen den Mittelwerten im Vergleich Stadtkinder vs. Landkinder.

Wie in Tabelle 26 ersichtlich existiert ein deskriptiver Unterschied zwischen Stadt- und Landkindern bezüglich des On-Task-Verhaltens. Hier zeigen sich die Stadtkinder etwas engagierter als die Landkinder. Allerdings werden die Unterschiede statistisch nicht signifikant (Lehrer: $F [1, 207] = 3.41, p = .060$; Trainer: $F [1, 207] = 3.72, p = .060$). Ein kleiner signifikanter Unterschied zwischen Stadt- und Landkindern findet sich jedoch beim Off-Task-Verhalten. Allerdings nur in den Lehrereinschätzung (Lehrer: $F [1, 207] = 4.09, p = .040$). Die Landkinder werden von ihren Lehrern als etwas störender bewertet als die Stadtkinder. Bei den Trainereinschätzungen zeigen sich hier jedoch keine Unterschiede (Trainer:

$F [1, 207] = 0.17, p = .680$). Hinzu kommt, dass die notwendigen Voraussetzungen der hier verwendeten multivariaten Varianzanalyse für die Lehrereinschätzungen des Off-Task-Verhaltens nicht alle erfüllt sind. Die Homogenität der Fehlervarianzen und Kovarianzen ist laut Levene-Test ($F [1, 207] = 5.70, p = .020$) und Box-M-Test ($F [21, 28401] = 2.08, p = .000$) nicht gegeben. Da außerdem die Gruppengrößen von Stadtkindern ($n = 94$) und Landkindern ($n = 308$) sehr unterschiedlich sind, entscheidet der F-Test hier progressiver. Dies muss bei der Interpretation des Ergebnisses beachtet werden, weshalb dieser kleine Unterschied im Off-Task-Verhalten nicht überbewertet werden sollte.

Der Spaß der Kinder am Programm wird bei Stadt- und Landkindern in etwa gleich hoch eingeschätzt. Ein ähnliches Muster der Mittelwerte findet sich auch, wenn die drei verschiedenen Interventionsbereiche einzeln betrachtet werden.

Intelligenzeffekte (Ravenmatrizen)

Auch die über die Ravenmatrizen ermittelte Intelligenz der Kinder hat einen leichten Einfluss auf die Implementation von PARTS. Dieser Einfluss zeigt sich allerdings ausschließlich beim Off-Task-Verhalten der Kinder. Die Korrelationen der Intelligenz mit den Implementationsparametern werden in Tabelle 27 wiedergegeben.

Tabelle 27

Korrelationen zwischen der Intelligenz der Kinder zum ersten Messzeitpunkt und den mittleren Implementationsparametern

	On-Task-Verhalten	Off-Task-Verhalten	Spaß
Lehrer	.07	-.15*	.08
Trainer	.03	-.19*	.03

Anmerkungen. * = signifikante Produkt-Moment-Korrelation nach Pearson mit $p < 0.05$.

Kinder, die bei den Ravenmatrizen mehr Punkte erzielen konnten, zeigten weniger Off-Task-Verhalten, als Kinder mit einem niedrigeren Ergebnis im Matrizen-Test. Dies wird sowohl in der Lehrer- als auch in der Trainereinschätzung statistisch bedeutsam, auch wenn die Stärke des Zusammenhangs eher gering ist. Zu den anderen beiden Parametern On-Task-Verhalten und Spaß finden sich keine Zusammenhänge mit dem Ravenmatrizentest. Für die einzelnen

Interventionsbereiche ergibt sich ein ähnliches Korrelationsmuster wie für das Gesamtprogramm.

Implementation bei Störenfrieden vs. Nicht-Störenfriede

Die Kinder, welche über mindestens die Hälfte aller Interventionssitzungen hinweg ein deutlich erhöhtes Off-Task-Verhalten gezeigt haben (s.o., Störenfriede, $n = 26$), unterscheiden sich auch auf den anderen Implementationsparametern von Kindern mit weniger starkem Off-Task-Verhalten (Nicht-Störenfriede, $n = 183$).

Tabelle 28 zeigt die durchschnittlichen Implementationsparameter dieser beiden Gruppen im Vergleich, getrennt nach Lehrer- und Trainereinschätzungen.

Tabelle 28

Mittlere Implementationsparameter für Störenfriede versus Nicht-Störenfriede, getrennt nach Lehrer- und Trainereinschätzungen

	On-Task-Verhalten		Off-Task-Verhalten		Spaß	
	ST	N-ST	ST	N-ST	ST	N-ST
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Lehrer	1.85	2.16	1.40	0.32	1.47	1.57
	(0.58)	(0.49)	(0.50)	(0.36)	(0.38)	(0.32)
	$d = -0.58^*$		$d = 2.48^*$		$d = -0.29$	
Trainer	1.84	2.02	1.09	0.22	1.38	1.50
	(0.52)	(0.54)	(0.43)	(0.24)	(0.44)	(0.28)
	$d = -0.34$		$d = 2.50^*$		$d = -0.33$	

Anmerkungen. ST = Störenfriede; N-ST = Nicht-Störenfriede; d = Effektstärke im Vergleich Störenfriede vs. Nicht-Störenfriede; * = signifikanter Unterschied mit $p < 0.05$ zwischen den Mittelwerten im Vergleich Störenfriede vs. Nicht-Störenfriede.

Wie anhand der Mittelwerte und Effektstärken ersichtlich ist, finden sich, abgesehen vom Off-Task-Verhalten nur kleine bis mittelgroße Unterschiede zwischen den beiden Gruppen. Die Nicht-Störenfriede haben ein etwas besseres On-Task-Verhalten gezeigt als die Störenfriede, wobei der Mittelwertsunterschied lediglich bei den Lehrereinschätzungen statistisch bedeutsam wird (Lehrer: $F [1, 207] = 8.69, p = .004$; Trainer: $F [1, 207] = 2.65, p = .105$). Auch dieser

Unterschied sollte mit Vorsicht interpretiert werden, da bei der hierzu verwendeten MANOVA der Box-M-Test auf Gleichheit der Kovarianzen der Fehler signifikant wurde ($F [21, 6996.39] = 4.12, p = .000$), und die Gruppengrößen deutlich unterschiedlich sind – der F-Test hier also progressiv reagiert.

Die deutlichen Mittelwertsunterschiede im Off-Task-Verhalten der beiden Gruppen sind zu erwarten, da sie sich aus dem Auswahlkriterium der beiden Gruppen ergeben. Diese Unterschiede sind auch statistisch signifikant (Lehrer: $F [1, 207] = 26.32, p = .000$; Trainer: $F [1, 207] = 17.33, p = .000$).

Beim dritten Implementationsparameter „Spaß am Programm“ sind die Unterschiede nur klein und werden nicht statistisch signifikant (Lehrer: $F [1, 207] = 0.27, p = .115$; Trainer: $F [1, 207] = 0.36, p = .053$), weisen aber in die erwartete Richtung. Störenfriede zeigen etwas weniger Spaß am Programm als Nicht-Störenfriede.

IV.3.5. Implementationsanalyse – Zusammenfassung

Die Analysen zur Implementation von PARTS beruhen auf den Protokollen der einzelnen Interventionseinheiten von begleitenden Klassenlehrern und durchführenden Trainern. Die auf Basis dieser Daten oben aufgeführten Analysen zur Implementation von PARTS zeichnen insgesamt ein sehr positives Bild. Die Kinder zeigen über das gesamte Programm hinweg ein – verglichen mit dem Wertebereich der verwendeten Skalen – überdurchschnittlich hohes On-Task-Verhalten, ein sehr niedriges Off-Task-Verhalten und sehr viel Spaß an den einzelnen Sitzungen.

Dabei sticht aus den einzelnen Interventionsbereichen wissensbasierte Intervention, Stellvertretender-Kontakt und Förderung sozial-kognitiver Fähigkeiten der Bereich der wissensbasierten Intervention, vor allem den Spaß der teilnehmenden Kinder betreffend, besonders positiv hervor. Auch die anderen Bereiche ließen sich sehr gut implementieren; nur der Bereich Förderung der sozial-kognitiven Fähigkeiten begeisterte die Kinder etwas weniger. Hier zeigen sich beim On-Task-Verhalten und beim Spaß signifikante Unterschiede zu den übrigen Interventionsbereichen.

Eine kleine Gruppe von Kindern mit dauerhaft erhöhtem Off-Task-Verhalten („Störenfriede“, $n = 26$) konnte identifiziert werden. Diese Gruppe unterschied sich, wie zu erwarten, auch in den Bereichen On-Task-Verhalten und Spaß von der

Gruppe der übrigen Teilnehmer. Allerdings blieben die Unterschiede auf einem geringen bis mittleren Niveau.

Bei den möglichen Einflussfaktoren auf die Implementation zeigen sich Effekte der Kovariaten Geschlecht, Alter, Entwicklungskontext und Intelligenz. Bei Mädchen konnte das Programm etwas besser als bei Jungen implementiert werden. Jüngere Kinder zeigten etwas mehr Beteiligung am Programm als ältere, ebenso verhält es sich bei den Stadtkindern im Vergleich zu den Landkindern. Kinder, die ein besseres Ergebnis im Standardmatrizentest erreichten, zeigten sich etwas weniger unaufmerksam und störend während des Programms. Diese Effekte der Einflussfaktoren bleiben aber durchweg auf einem kleinen bis maximal mittleren Niveau, erreichen aber meist das statistische Signifikanzniveau.

Das Präventionsprogramm PARTS konnte also insgesamt sehr gut implementiert werden. Dafür sprechen sowohl die Ergebnisse der obigen Implementationsanalyse mit den drei verwendeten Implementationsparametern On-Task-Verhalten, Off-Task-Verhalten und Spaß, als auch die persönlichen Erfahrungen der Trainer, die von den teilnehmenden Kindern zu den einzelnen Sitzungen meist begeistert begrüßt wurden und viel positives Feedback, sowohl von den Kindern als auch von den beteiligten Lehrern, erhielten.

IV.4. Wirksamkeitsanalyse

Im Folgenden wird die Wirksamkeit des Präventionsprogramms PARTS auf die proximalen und distalen Effektmaße geprüft. Dazu finden Vergleiche der Entwicklung von Interventions- und Kontrollgruppe über die Zeit statt. Dabei handelt es sich sowohl um Prä-Post-Vergleiche zur Prüfung direkter kurzfristiger Effekte, als auch um Prä-Follow-up-Vergleiche zur Aufdeckung von verzögerten oder langfristigen Effekten von PARTS. Zum einen erfolgen diese Vergleiche deskriptiv anhand von Mittelwerten und Effektstärken. Zum anderen wird eine inferenzstatistische Prüfung der vorgefundenen Mittelwertsunterschiede via Varianzanalysen mit und ohne Messwiederholungen¹⁹ durchgeführt. Der Einfluss der möglichen Kovariaten (Geschlecht, Alter, Entwicklungskontext, Intelligenz) auf die Evaluationsergebnisse wird ebenfalls mit Hilfe der entsprechenden Kovarianzanalysen geprüft. Ergebnisse zu den Kovariaten werden allerdings nur berichtet, falls diese bedeutsamen Einfluss auf die globalen Evaluationsergebnisse haben. Da sich bei keiner der Auswertungen der Wirksamkeitsanalyse ein statistisch bedeutsamer Einfluss der Kovariaten auf die Evaluationsergebnisse zeigte, werden die Angaben der entsprechenden Kovarianzanalysen nicht berichtet. Die deskriptiven und inferenzstatistischen Ergebnisse für die Risikogruppe 25 der jeweiligen Skala werden im Anhang D dargestellt.

Zunächst werden die Ergebnisse zu den proximalen Effektmaßen berichtet. Im Anschluss daran die Ergebnisse der distalen Effektmaße.

IV.4.1. Ergebnisse zu den proximalen Effektmaßen

Die proximalen Effektmaße sind die Variablen, welche direkt durch das Programm beeinflusst werden sollen. Dazu gehören die Variablen Interkulturelles Wissen, Multiples Klassifizieren, Perspektivenübernahmefähigkeit und die soziale Informationsverarbeitung. Die Evaluationsergebnisse zu diesen Maßen werden in den nächsten Abschnitten vorgestellt.

¹⁹ Abhängig von der vorliegenden Datenstruktur werden Varianzanalysen mit Messwiederholungen verwendet, wenn Daten der Variablen über alle drei Messzeitpunkte vorliegen. Fehlen Daten des ersten Messzeitpunktes, werden Varianzanalysen ohne Messwiederholungen zur Ermittlung der Mittelwertsunterschiede innerhalb eines Messzeitpunktes durchgeführt.

Interkulturelles Wissen

Die Ergebnisse der Teilnehmer im Wissenstest liegen alle im Bereich zwischen 1 und 23 Punkten und reizen damit den Wertebereich (0-24 Punkte) fast vollständig aus. Dabei unterscheidet sich das kulturelle Wissen der Kinder zum ersten Messzeitpunkt nicht bedeutsam zwischen Interventions- und Kontrollgruppe ($F [1, 397] = 2.61, p = .107$). In der Interventionsgruppe nimmt das kulturelle Wissen über die Zeit deutlich stärker zu, als in der Kontrollgruppe. Dies schlägt sich auch in den großen Nettoeffektstärken ($d_{t2-t1} = 0.78$ und $d_{t3-t1} = 1.14$) nieder. Die Mittelwerte und Standardabweichungen sowie die Netto-Effektstärken zum interkulturellen Wissen lassen sich in Tabelle 29 ablesen.

Tabelle 29

Mittelwerte, Standardabweichungen und Netto-Effektstärken des interkulturellen Wissens getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
Interkulturelles Wissen	IG	9.91 (3.24)	13.41 (3.56)	15.26 (3.81)	0.78*	1.14*
	KG	9.39 (3.22)	10.38 (3.15)	11.05 (3.31)		

Anmerkungen: IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß; * = signifikant unterschiedliche Entwicklungen der Untersuchungsgruppen mit $p < 0.05$.

Eine Varianzanalyse mit Messwiederholungen bestätigt die statistische Bedeutsamkeit dieser Gruppenunterschiede in der Entwicklung mit einer signifikanten Interaktion der Faktoren Messzeitpunkt und Gruppe ($F [2, 396] = 51.98, p = .000$). Dabei zeigen die Post-hoc-Kontraste, dass sowohl die Entwicklung vom ersten zum zweiten Messzeitpunkt zwischen den Gruppen unterschiedlich verläuft ($F [1, 397] = 51.55, p = .000$), als auch die Entwicklung vom zweiten zum dritten Messzeitpunkt ($F [1, 397] = 64.11, p = .000$), ebenso wie die Entwicklung vom ersten zum dritten Messzeitpunkt ($F [1, 397] = 100.93, p = .000$).

Zusätzlich finden sich signifikante Haupteffekte der Faktoren Gruppe ($F [1, 397] = 87.80, p = .000$) und Messzeitpunkte ($F [2, 396] = 158.94, p = .000$).

Multiple Klassifikationsfähigkeit

Die Werte der Teilnehmer beim Klassifikationstest mit sozialen Stimuli decken den gesamten Wertebereich der Skala ab und liegen zwischen 0 und 12. Wie in Tabelle 30 zu sehen ist, steigen in beiden Untersuchungsgruppen die Werte des Klassifikationstests mit sozialen Stimuli über die Zeit an. Dabei unterscheiden sich die Ausgangswerte der Gruppen nicht bedeutsam voneinander ($F [1, 398] = 0.15$, $p = .698$). Der Zuwachs in der Interventionsgruppe ist minimal größer als in der Kontrollgruppe, was sich in kleinen positiven Effektstärken widerspiegelt.

Tabelle 30

Mittelwerte, Standardabweichungen und Netto-Effektstärken der multiplen Klassifikationsfähigkeit mit sozialen Stimuli, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		M (SD)	M (SD)	M (SD)		
Multiple Klassifikation	IG	7.64 (2.45)	9.01 (2.39)	9.54 (2.29)	0.13	0.16
	KG	7.74 (2.46)	8.78 (2.34)	9.24 (2.38)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

Inferenzstatistisch wird dieser Effekt jedoch nicht signifikant. Es ergibt sich keine statistisch bedeutsame Interaktion der Faktoren Gruppe und Messzeitpunkt ($F [2, 394] = 0.77$, $p = .466$). Auch der Haupteffekt des Faktors Gruppe wird hier nicht signifikant ($F [1, 395] = 0.77$, $p = .381$). Allein der Faktor Messzeitpunkt hat einen bedeutsamen Haupteffekt auf die multiple Klassifikationsfähigkeit ($F [2, 394] = 62.82$, $p = .000$). Dementsprechend steigt also die multiple Klassifikationsfähigkeit in beiden Gruppen über die Zeit signifikant an.

Perspektivenübernahme

Die Mittelwerte der Perspektivenübernahmefähigkeit liegen bereits zum ersten Messzeitpunkt auf einem relativ hohen Niveau zwischen $M_{IG} = 3.02$ und $M_{KG} = 3.13$

bei einem voll ausgenutzten Wertebereich der Skala von null bis vier. Dabei unterscheiden sich die Mittelwerte der beiden Untersuchungsgruppen nicht signifikant voneinander ($F [1, 394] = 1.34, p = .249$). Wie in Tabelle 31 zu sehen ist, steigt dieser Wert in der Interventionsgruppe über die Zeit leicht an. In der Kontrollgruppe hingegen bleiben die Werte zwischen Prä- und Posttest auf dem gleichen Level und steigen nur zum Follow-up hin leicht an. Diese unterschiedlichen Entwicklungsverläufe machen sich auch in den kleinen positiven Netto-Effektstärken bemerkbar.

Tabelle 31

Mittelwerte, Standardabweichungen und Netto-Effektstärken der Perspektivenübernahmefähigkeit, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		M (SD)	M (SD)	M (SD)		
Perspektivenübernahme	IG	3.02 (0.99)	3.17 (0.94)	3.38 (0.83)	0.18	0.20
	KG	3.13 (0.94)	3.11 (0.95)	3.30 (0.87)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

Es ergibt sich keine statistisch bedeutsame Interaktion der Faktoren Gruppe und Messzeitpunkt ($F [2, 392] = 1.37, p = .256$). Auch der Haupteffekt des Faktors Gruppe wird hier nicht signifikant ($F [1, 393] = 0.03, p = .870$). Allein der Faktor Messzeitpunkte hat einen bedeutsamen Haupteffekt auf die Perspektivenübernahmefähigkeit ($F [2, 392] = 11.56, p = .000$). In beiden Gruppen steigt die Fähigkeit zur Perspektivenübernahme über die Zeit (zumindest zum Follow-up hin) signifikant an.

Soziale Informationsverarbeitung

Der Wertebereich des Tests zum sozialen Problemlösen liegt zwischen 6 und 15 und wird von den Teilnehmern der Untersuchung voll ausgeschöpft. Die

vorgefundenen Mittelwerte liegen in beiden Gruppen über die Zeit hinweg stabil auf einem relativ hohen Wertenniveau zwischen 11 und 12 Punkten. Die Mittelwerte des Prätests unterscheiden sich nicht signifikant zwischen den Untersuchungsgruppen ($F [1, 397] = 2.04, p = .155$). Dabei zeigen sich weder bei der Interventionsgruppe noch bei der Kontrollgruppe deutliche Veränderungen über die Zeit. Dies spiegelt sich auch in den Netto-Effektstärken wider, welche sich nahe Null bewegen. Die Werte sind in Tabelle 32 dargestellt.

Tabelle 32

Mittelwerte, Standardabweichungen und Netto-Effektstärken des Tests zum sozialen Problemlösen, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
Soziale Informationsverarbeitung	IG	11.59 (1.42)	11.55 (1.58)	11.69 (1.42)	-0.04	-0.03
	KG	11.38 (1.58)	11.40 (1.36)	11.52 (1.33)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

Es ergibt sich keine statistisch bedeutsame Interaktion der Faktoren Gruppe und Messzeitpunkt ($F [2, 395] = 0.09, p = .917$). Auch die Haupteffekte der Faktoren Gruppe ($F [1, 396] = 3.31, p = .070$) und Messzeitpunkt ($F [2,395] = 1.73, p = .178$) werden hier nicht signifikant.

Zusammenfassung der Ergebnisse zu den proximalen Effektmaßen

Auf einem Großteil der proximalen Effektmaße findet sich insgesamt ein Zuwachs der gemessenen Werte. So zeigt sich beispielsweise im Bereich des interkulturellen Wissens bei allen Kindern ein deutlicher Wissenszuwachs über die Zeit, der durch das Interventionsprogramm jedoch deutlich gesteigert wird. Die Interventionsgruppe konnte dementsprechend einen signifikant größeren Wissenszuwachs verzeichnen als die Kontrollgruppe. Dies gilt für beide

Vergleichszeiträume und insbesondere auch für die Risikogruppe im interkulturellen Wissen (s. Anhang D).

Auch im Bereich der multiplen Klassifikationsfähigkeit zeigt sich eine positive Entwicklung beider Untersuchungsgruppen über die Zeit. Hier findet sich nur ein kleiner nicht-signifikanter Vorsprung der Interventionsgruppe gegenüber der Kontrollgruppe. Die Risikogruppe 25 zeigt ebenfalls eine positive Entwicklung im Bereich der multiplen Klassifikation, wobei die Interventionsgruppe hier einen deutlich größeren Zuwachs zeigt als die Kontrollgruppe. Der Effekt wird aber statistisch nicht signifikant (s. Anhang D).

Im Bereich der Perspektivenübernahme findet sich eine leichte positive Entwicklung in beiden Untersuchungsgruppen, wobei die Interventionsgruppe ihre Fähigkeiten über die Zeit etwas mehr steigern kann als die Kontrollgruppe. Bei der Risikogruppe 25 zeigt sich erst verzögert zum Follow-up hin eine positivere Entwicklung der Interventionsgruppe im Vergleich zur Kontrollgruppe. Auch hier lassen sich die leicht unterschiedlichen Entwicklungsverläufe nicht statistisch absichern (s. Anhang D).

Die soziale Informationsverarbeitung bleibt über die Zeit hinweg stabil auf dem gleichen relativ hohen Niveau. Die Risikogruppe 25 erreicht über die Zeit in beiden Untersuchungsgruppen fast das Niveau der Gesamtgruppe, wobei die soziale Informationsverarbeitung der Interventionsgruppe sich zum Posttest hin stärker verbessert als die der Kontrollgruppe (s. Anhang D).

Insgesamt findet also, abgesehen von der sozialen Informationsverarbeitung ein Zuwachs der gemessenen Fähigkeiten auf den proximalen Effektmaßen statt. Die Interventionsgruppe kann gegenüber der Kontrollgruppe, insbesondere beim interkulturellen Wissen, signifikant, aber auch im Bereich der multiplen Klassifikation und der Perspektivenübernahmefähigkeit, zumindest tendentiell, vom Interventionsprogramm profitieren. Keinerlei Effekte finden sich auf der Skala zum sozialen Problemlösen.

IV.4.2. Ergebnisse zu den distalen Effektmaßen

Die distalen Effektmaße sind die Variablen, welche die eigentlichen Zielkriterien von PARTS abbilden, die aber nicht direkt durch das Programm trainiert werden. Darunter finden sich gruppenbezogene soziale Einstellungsmaße mit kognitiven (via Attributszuordnung: MRA und Verkuyten), affektiven (via Sympathie-Einschätzungen: IBM und Verkyuten, oder soziale Distanz: Haus-Aufgabe) und konativen (via Verhaltensabsichten zur Kontaktaufnahme: IBM) Schwerpunkten. Weiterhin gehören allgemeine interpersonale Einstellungen (Toleranz) und das Sozialverhalten der Teilnehmer (via Eltern- und Lehrereinschätzungen: SDQ) zu den distalen Effektmaßen. Die Ergebnisse zu diesen Maßen werden im folgenden Abschnitt dargestellt.

IV.4.3. Gruppenbezogene soziale Einstellungen mit kognitivem Schwerpunkt

Russische Zielfremdgruppe

Personenzentrierte Attributszuordnung – Multi-Response Racial Attitude Measure

Eigengruppenbevorzugung. Die Werte der MRA-Eigengruppenbevorzugung gegenüber einem russischem Kind verteilen sich über einen Wertebereich von -17 bis 21 bei einer Gesamtskala von -21 bis 21. Die Mittelwerte liegen zwischen $M = 1.44$ und $M = 2.39$, und sind damit leicht höher als bei der Eigengruppenbevorzugung gegenüber einem chinesischen Kind (s.u.), aber insgesamt noch sehr niedrig. Dabei unterscheiden sich die Prätestmittelwerte der beiden Untersuchungsgruppen nicht bedeutsam voneinander ($F [1, 398] = 1.14$, $p = .286$).

Die Entwicklungsverläufe der Untersuchungsgruppen unterscheiden sich jedoch leicht voneinander. Die Interventionsgruppe erfährt nach dem Prätest zunächst einen kleinen Anstieg in ihrer Eigengruppenbevorzugung, bevor diese zum Follow-up hin wieder etwas sinkt. In der Kontrollgruppe sieht der Entwicklungsverlauf umgekehrt aus. Zwischen Prä- und Posttest ergibt sich ein Abfall in der Eigengruppenbevorzugung, während es zum Follow-up hin wieder einen Anstieg zu verzeichnen gibt.

Diese unterschiedlichen Entwicklungsverläufe spiegeln sich auch in den kleinen Netto-Effektstärken wider, die im Prä-Post-Vergleich noch positiv ausfällt, im Prä-Follow-up-Vergleich hingegen negativ wird. Dies wird aus Tabelle 33 deutlich.

Tabelle 33

Mittelwerte, Standardabweichungen und Netto-Effektstärken der MRA-Eigengruppenbevorzugung (MRA-EGB) gegenüber einem russischem Kind (r), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
MRA-EGB (r)	IG	1.72	2.15	1.44	0.19	-0.13
		(4.21)	(4.75)	(4.58)		
	KG	2.16	1.81	2.39		
		(3.91)	(4.45)	(4.28)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

Diese gegenläufige Entwicklung in den beiden Untersuchungsgruppen wird auch statistisch bedeutsam. So findet sich hier eine signifikante Interaktion der Faktoren Messzeitpunkte und Gruppe ($F [2, 395] = 4.03, p = .018$). Bei Betrachtung der Post-hoc-Kontraste wird jedoch weder der Entwicklungsunterschied zwischen erstem und zweitem Messzeitpunkt ($F [1, 396] = 1.59, p = .208$), noch der Unterschied zwischen erstem und drittem Messzeitpunkt ($F [1, 396] = 1.43, p = .232$) statistisch bedeutsam. Allein der Kontrast zwischen dem zweiten und dem dritten Messzeitpunkt zeigt eine signifikante Interaktion der Faktoren Messzeitpunkte und Gruppe ($F [1, 396] = 5.63, p = .018$). Ansonsten werden weder der Haupteffekt für den Faktor Messzeitpunkte ($F [2, 395] = 0.08, p = .925$) noch der Haupteffekt des Faktors Gruppe ($F [1, 396] = 1.26, p = .263$) statistisch signifikant.

Fremdgruppenabwertung. Die Werte der MRA-Fremdgruppenabwertung gegenüber einem russischen Kind liegen zwischen -17 und 21 und nutzen die Gesamtskala von -21 bis 21 damit fast vollkommen aus. Die Mittelwerte sind dabei jedoch relativ klein und bleiben nahe an Null. Dabei unterscheiden sich die

Prätestmittel zwischen den beiden Untersuchungsgruppen nicht signifikant voneinander ($F [1, 398] = 0.26, p = .610$).

Die Entwicklungsverläufe der beiden Untersuchungsgruppen unterscheiden sich allerdings. Während die Kontrollgruppe über alle drei Messzeitpunkte relativ stabil in ihrer Fremdgruppenabwertung gegenüber einem russischen Kind bleibt, verändert sich das Niveau der Fremdgruppenabwertung in der Interventionsgruppe schon, zumindest zwischen dem zweiten und dritten Messzeitpunkt. Hier findet ein Abfall der Fremdgruppenabwertung in der Interventionsgruppe statt. Dieser Unterschied in den Entwicklungsverläufen macht sich auch in der kleinen negativen Netto-Effektstärke im Prä-Follow-up-Vergleich bemerkbar, während die Netto-Effektstärke im Prä-Post-Vergleich kaum von Null verschieden ist. Mittelwerte, Standardabweichungen und Netto-Effektstärken lassen sich in Tabelle 34 ablesen.

Tabelle 34

Mittelwerte, Standardabweichungen und Netto-Effektstärken der MRA-Fremdgruppenabwertung (MRA-FGA) gegenüber einem russischen Kind (r), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3		
		<i>M</i>	<i>M</i>	<i>M</i>	d_{t2-t1}	d_{t3-t1}
		(SD)	(SD)	(SD)		
MRA_FGA (r)	IG	1.91	2.18	1.07	0.08	-0.21
		(5.09)	(5.10)	(4.56)		
	KG	2.15	2.07	2.26		
		(4.12)	(4.55)	(4.93)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

Die oben beschriebenen unterschiedlichen Entwicklungsverläufe lassen sich auch inferenzstatistisch stützen. So findet sich eine bedeutsame Interaktion der Faktoren Messzeitpunkte und Gruppe ($F [2, 395] = 3.16, p = .044$). Allerdings zeigen sich bei den Post-hoc-Kontrasten weder zwischen dem ersten und dem zweiten Messzeitpunkt ($F [1, 396] = 0.26, p = .614$), noch zwischen dem ersten und dem dritten Messzeitpunkt ($F [1, 396] = 2.78, p = .096$) signifikante Kontraste. Auch hier wird allein der Kontrast zwischen dem zweiten und dem dritten Messzeitpunkt

statistisch bedeutsam ($F [1, 396] = 5.71, p = .017$). Ansonsten werden weder der Haupteffekt des Faktors Gruppe ($F [1, 396] = 1.70, p = .194$), noch der Haupteffekt des Faktors Messzeitpunkte statistisch bedeutsam ($F [2, 395] = 1.74, p = .177$).

Gruppenzentrierte Attributszuordnung – Verkuyten-Skala

Die Verkuyten-Skala wurde nur zum dritten Messzeitpunkt auch mit russischen Kindern als Zielfremdgruppe verwendet. Die Werte verteilen sich hier zwischen -1.33 bis 2.67 bei einem Wertebereich der Skala von -3 bis 3. Die Mittelwerte von Interventions- und Kontrollgruppe liegen relativ nahe dem Nullpunkt, wobei die Teilnehmer der Kontrollgruppe eine etwas größere Einstellungsverzerrung gegenüber der Gruppe der russischen Kinder zeigen, als die Kinder der Interventionsgruppe. Dies zeigt sich auch an der kleinen negativen Effektstärke, wie in Tabelle 35 zu sehen ist.

Tabelle 35

Mittelwerte, Standardabweichungen und Netto-Effektstärken der kognitiven Verkuyten-Skala gegenüber russischen Kindern (r), getrennt nach Interventionsgruppe und Kontrollgruppe zum dritten Messzeitpunkt

		T3	
		<i>M (SD)</i>	<i>d_{t3}</i>
Verkuyten-Skala (r)	IG	0.32 (0.59)	-0.29*
	KG	0.51 (0.71)	

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; *d* = Effektstärkemaß; * = signifikanter Mittelwertsunterschied mit $p < 0.05$.

Der Mittelwertsunterschied zwischen Interventions- und Kontrollgruppe wird bei der Varianzanalyse auch inferenzstatistisch signifikant ($F [1, 400] = 9.29, p = .002$).

Chinesische Fremdgruppe

Personenzentrierte Attributszuordnung – Multi-Response Racial Attitude Measure

Eigengruppenbevorzugung. Die MRA-Eigengruppenbevorzugung gegenüber einem chinesischen Kind liegen im Mittel nahe Null. Dabei verteilen sich die Werte der einzelnen Teilnehmer von -19 bis 21 über fast den gesamten Wertebereich (-21 bis 21).

Die Eigengruppenbevorzugung der Teilnehmer der Interventionsgruppe nimmt dabei über die Zeit ab. Auch in der Kontrollgruppe findet zunächst eine deutliche Abnahme der Eigengruppenbevorzugung statt. Allerdings steigt der Wert in der Kontrollgruppe vom zweiten zum dritten Messzeitpunkt minimal wieder an.

Die Ausgangsniveaus der beiden Untersuchungsgruppen unterscheiden sich relativ stark, aber nicht signifikant voneinander ($F [1, 398] = 3.29, p = .070$). Die Kontrollgruppe startet auf einem höheren Biasniveau als die Interventionsgruppe. Zum zweiten Messzeitpunkt gleichen sich diese Niveaus jedoch stark an. Zum dritten Messzeitpunkt reduziert sich die Eigengruppenbevorzugung in der Interventionsgruppe weiter und reicht sogar in den negativen Wertebereich hinein, während die Eigengruppenbevorzugung der Kontrollgruppe wieder ganz leicht ansteigt. Dies erzeugt die kleinen positiven Netto-Effektstärken im Prä-Post- und im Prä-Follow-up-Vergleich, wie in Tabelle 36 deutlich wird.

Tabelle 36

Mittelwerte, Standardabweichungen und Netto-Effektstärken der MRA-

Eingengruppenbevorzugung (MRA-EGB) gegenüber einem chinesischen Kind (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
MRA-EGB (c)	IG	0.18	0.06	-0.09	0.18	0.13
		(5.16)	(5.42)	(5.19)		
	KG	1.04	0.09	0.17		
		(4.22)	(4.84)	(4.75)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

In der inferenzstatistischen Analyse werden die oben beschriebenen Entwicklungen und Entwicklungsunterschiede allerdings nicht bedeutsam. Die multivariaten Tests der Varianzanalyse mit Messwiederholung zeigen weder eine signifikante Interaktion der Faktoren Gruppe und Messzeitpunkte ($F [2, 395] = 0.78$, $p = .460$), noch einen signifikanten Haupteffekt der Faktoren Gruppe ($F [1, 396] = 1.18$, $p = .279$) oder Messzeitpunkte ($F [2, 395] = 2.26$, $p = .106$).

Fremdgruppenabwertung. Die MRA-Fremdgruppenabwertung gegenüber einem chinesischen Kind erstreckt sich über einen Wertebereich von -17 bis 21 und schöpft damit fast den gesamten Wertebereich dieser Skala aus (-21 bis 21). Die Mittelwerte der Skala liegen jedoch nahe Null. Die Prätestwerte der Skala unterscheiden sich dabei nicht bedeutsam zwischen den Untersuchungsgruppen ($F [1, 398] = 1.35$, $p = .246$).

Es findet in beiden Untersuchungsgruppen ein leichter Abfall der Fremdgruppenabwertung über die Zeit statt. Dieser Abfall ist jedoch sehr gering und in den beiden Gruppen nahezu parallel, so dass die Netto-Effektstärken ebenfalls fast gleich Null sind. Diese Entwicklung lässt sich aus den Mittelwerten und Netto-Effektstärken aus Tabelle 37 ablesen.

Tabelle 37

Mittelwerte, Standardabweichungen und Netto-Effektstärken der MRA-Fremdgruppenabwertung (MRA-FGA) gegenüber einem chinesischen Kind (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		M (SD)	M (SD)	M (SD)		
MRA_FGA (c)	IG	0.28 (4.34)	0.01 (3.88)	-0.37 (4.74)	0.09	0.01
	KG	0.77 (4.01)	0.14 (3.35)	0.07 (3.65)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

Die oben beschriebene Entwicklung spiegelt sich auch in der inferenzstatistischen Analyse wider. Es finden sich weder eine signifikante Interaktion der Faktoren Gruppe und Messzeitpunkte ($F [2, 395] = 0.31, p = .732$), noch ein bedeutsamer Haupteffekt des Faktors Gruppe ($F [1, 396] = 1.48, p = .225$). Allein der Faktor Messzeitpunkte zeigt einen signifikanten Haupteffekt ($F [2, 395] = 4.07, p = .036$).

Gruppenzentrierte Attributszuordnung – Verkuyten-Skala

Die Biaswerte der Verkuyten-Skala gegenüber chinesischen Kindern bewegen sich in einem Wertebereich von -2.00 bis zu 3.00 bei einer Skala von -3.00 bis 3.00, wobei negative Werte eine Bevorzugung der Fremdgruppe und positive Werte eine Bevorzugung der Eigengruppe anzeigen. Die Mittelwerte in den beiden Untersuchungsgruppen sind positiv und liegen nahe dem Nullpunkt, was nur eine leichte Bevorzugung der Eigengruppe anzeigt. Die Prätestwerte der beiden Untersuchungsgruppen unterscheiden sich allerdings signifikant voneinander ($F [1, 398] = 6.67, p = .010$), wobei der Wert in der Interventionsgruppe niedriger ist als in der Kontrollgruppe. In beiden Gruppen findet sich ein Absinken der Biaswerte von t1 zu t2. Danach bleiben die Werte stabil bzw. steigen bei der Interventionsgruppe wieder leicht an. Allerdings liegen die Werte bei der Interventionsgruppe insgesamt auf einem leicht niedrigeren Niveau, als die Werte in der Kontrollgruppe. Die beiden Untersuchungsgruppen unterscheiden sich kaum in ihrer Entwicklung, was sich auch in den nahe Null liegenden Netto-Effektstärken in Tabelle 38 ablesen lässt.

Tabelle 38

Mittelwerte, Standardabweichungen und Netto-Effektstärken der kognitiven Verkuyten-Skala gegenüber chinesischen Kindern (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
Verkuyten-Skala (c)	IG	0.26 (0.49)	0.16 (0.54)	0.21 (0.60)	0.02	0.12
	KG	0.41 (0.69)	0.30 (0.59)	0.29 (0.64)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

Da sich auf dieser Skala signifikante Prätest-Unterschiede zwischen den beiden Untersuchungsgruppen finden, werden an dieser Stelle zwei Regressionen berechnet, die sowohl den Prätestwert, als auch die Interaktion des Prätests mit dem Faktor Gruppenzugehörigkeit als Regressoren neben dem Regressor Gruppe simultan berücksichtigen. Dies geschieht, um mögliche Verzerrungen durch die Prätest-Unterschiede bei der Beurteilung der Ergebnisse herauszurechnen.

In der Regression auf den Posttest der kognitiven Verkuyten-Skala gegenüber der Gruppe chinesischer Kinder erweisen sich in dem signifikanten Regressionsmodell ($F [3, 396] = 11.96, p = .000$) mit einem von $r^2 = 0.08$ sowohl der Regressor Gruppe ($\beta = .13, p = .021$), als auch der Regressor Prätest ($\beta = .50, p = 0.005$) als statistisch bedeutsam, nicht jedoch der Interaktionsterm der beiden Regressoren ($\beta = -.26, p = .147$).

In der zweiten Regression auf das Follow-up der kognitiven Verkuyten-Skala gegenüber der Gruppe chinesischer Kinder zeigt sich nur noch der Regressor Prätest ($\beta = .35, p = .047$) im signifikanten Regressionsmodell ($F [3, 396] = 7.37, p = .000$) mit einem $r^2 = 0.05$ signifikant. Die Regressoren Gruppe ($\beta = .06, p = .311$) und der Interaktionsterm ($\beta = -.14, p = .435$) zeigen keinen bedeutsamen Einfluss auf das Follow-up.

Zusammenfassung der Ergebnisse zu gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt

Die gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt zeichnen sich fast durchgängig durch einen Abfall der Einstellungsverzerrungen in der gesamten Stichprobe aus. Dabei liegen die Werte der Einstellungsverzerrungen auch zum ersten Messzeitpunkt verglichen mit dem Wertebereich der Skalen auf einem sehr niedrigen Niveau. Auffällig ist hierbei, dass die Einstellungsverzerrungen gegenüber der chinesischen Fremdgruppe insgesamt noch etwas geringer sind, als die Einstellungsverzerrungen gegenüber der russischen Fremdgruppe. Bei den einzelnen Skalen zu den gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt ließen sich allerdings unterschiedliche Trends und Entwicklungsverläufe feststellen.

Bei der MRA-Eigengruppenbevorzugung gegenüber einem russischen Kind finden sich signifikant unterschiedliche Entwicklungsverläufe der beiden Untersuchungsgruppen. In der Kontrollgruppe gibt es zunächst einen leichten Abfall der Werte und zum dritten Messzeitpunkt einen Wiederanstieg. In der Interventionsgruppe ist der Verlauf genau umgekehrt. Hier gibt es zum zweiten Messzeitpunkt einen leichten Anstieg, aber zum dritten Messzeitpunkt einen Abfall der Werte. In der entsprechenden Risikogruppe 25 gibt es einen bedeutsamen Abfall der Werte vom ersten zum zweiten Messzeitpunkt. Dann findet sich in der Kontrollgruppe aber ein Wiederanstieg der Einstellungsverzerrungen, während es in der Interventionsgruppe einen weiteren Abfall der Werte gibt (s. Anhang).

Die MRA-Fremdgruppenabwertung gegenüber einem russischen Kind zeigt ebenfalls statistisch bedeutsame unterschiedliche Entwicklungsverläufe beider Untersuchungsgruppen. In der Interventionsgruppe findet sich von t2 zu t3 ein leichter Anstieg, von t2 zu t3 aber ein deutlicher Abfall in der Fremdgruppenabwertung. Dagegen bleibt das Niveau der Fremdgruppenabwertung in der Kontrollgruppe eher stabil. In der Risikogruppe 25 ergibt sich ebenfalls ein bedeutsamer Abfall der Werte in beiden Untersuchungsgruppen. Doch während es in der Kontrollgruppe zum dritten Messzeitpunkt einen Wiederanstieg gibt, fallen die Werte in der Interventionsgruppe weiter ab. Auch hier unterscheiden sich die Entwicklungsverläufe signifikant voneinander (s. Anhang).

Passend zu diesen Befunden, zeigt sich auf der Verkuyten-Skala gegenüber der Gruppe russischer Kinder ebenfalls ein bedeutsamer Mittelwertsunterschied

zwischen den beiden Untersuchungsgruppen. Die Interventionsgruppe zeigt auf der Verkuyten-Skala zum dritten Messzeitpunkt signifikant weniger Einstellungsverzerrungen als die Kontrollgruppe.

Bei dem personenzentrierten MRA gegenüber einem chinesischen Kind findet sich, ähnlich wie bei der kognitiven Verkuyten-Skala (s.u.), ein signifikanter Abfall in der Eigengruppenbevorzugung in beiden Untersuchungsgruppen. Dieser ist in der Kontrollgruppe noch etwas stärker als in der Interventionsgruppe, wobei die Kontrollgruppe hier auf einem insgesamt leicht höheren Niveau liegt als die Interventionsgruppe. Bei der entsprechenden Risikogruppe 25 findet ebenfalls ein signifikantes Absinken in den Einstellungsverzerrungen statt. Allerdings ergibt sich in der Kontrollgruppe zum dritten Messzeitpunkt ein leichter Wiederanstieg in der Eigengruppenbevorzugung, während die Interventionsgruppe auch zum dritten Messzeitpunkt weiter in ihrer Eigengruppenbevorzugung gegenüber einem chinesischen Kind absinkt (s. Anhang D).

Eine ähnliche Entwicklung zeigt sich im Bereich der Fremdgruppenabwertung gegenüber einem chinesischen Kind. Auch hier findet ein leichter aber signifikanter Abfall der Werte über die Zeit statt. Allerdings befinden sich die Werte hier auf einem sehr niedrigen Niveau, wobei die Interventionsgruppe auf einem noch etwas geringeren Level liegt, als die Kontrollgruppe. Die Entwicklung der beiden Gruppen läuft hier aber fast parallel. In der Risikogruppe 25 hingegen entwickeln sich die beiden Gruppen tendenziell unterschiedlich. Wie bei der Risikogruppe zur Eigengruppenbevorzugung findet ein bedeutsamer Abfall der Werte von t2 zu t3 in beiden Gruppen statt. Allerdings gibt es in der Kontrollgruppe zum dritten Messzeitpunkt einen leichten Wiederanstieg, während es in der Interventionsgruppe ein weiteres Absinken der Werte gibt (s. Anhang D).

Auf der Verkuyten-Skala gegenüber der Gruppe chinesischer Kinder schließlich fand in beiden Untersuchungsgruppen ein deutlicher Abfall der Einstellungsverzerrungen statt. Dieser Abfall war in der Kontrollgruppe sogar minimal größer als in der Interventionsgruppe, wobei das Niveau der Einstellungsverzerrungen bei der Interventionsgruppe etwas niedriger liegt, als in der Kontrollgruppe.

Insgesamt kann festgehalten werden, dass es signifikante Unterschiede zwischen den Untersuchungsgruppen bei den gruppenbezogenen sozialen

Einstellungen mit kognitivem Schwerpunkt gibt. Diese zeigen sich besonders gegenüber der russischen Fremdgruppe und hier hauptsächlich zwischen dem zweiten und dritten Messzeitpunkt.

IV.4.4. Gruppenbezogene soziale Einstellungen mit affektivem Schwerpunkt

Russische Zielfremdgruppe

Personenzentrierte Sympathieeinschätzung – Intended Behavior Measure-Sympathie-Item

Das IBM-Sympathie-Item wurde nur zum dritten Messzeitpunkt auch mit einem russischen Kind als Mitglied einer Zielfremdgruppe verwendet. Die Werte verteilen sich hier zwischen -4 bis 4 und nutzen damit den Wertebereich voll aus. Die Mittelwerte von Interventions- und Kontrollgruppe sind klein aber positiv, was für eine leichte Eigengruppenbevorzugung gegenüber dem russischen Kind spricht. Dabei zeigen die Teilnehmer der Kontrollgruppe eine etwas größere Einstellungsverzerrung gegenüber der Gruppe der russischen Kinder, als die Kinder der Interventionsgruppe. Dies zeigt sich auch an der kleinen negativen Effektstärke, wie in Tabelle 39 zu sehen ist.

Tabelle 39

Mittelwerte, Standardabweichungen und Netto-Effektstärken des IBM-Sympathie-Items gegenüber einem russischen Kind (r), getrennt nach Interventionsgruppe und Kontrollgruppe zum dritten Messzeitpunkt

		T3	
		<i>M (SD)</i>	<i>d_{t3}</i>
IBM-Sympathie (r)	IG	0.27 (1.15)	-0.32*
	KG	0.63 (1.10)	

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; *d* = Effektstärkemaß; * = signifikanter Mittelwertsunterschied mit $p < 0.05$.

Der Mittelwertsunterschied zwischen Interventions- und Kontrollgruppe wird bei der Varianzanalyse auch inferenzstatistisch signifikant ($F [1, 400] = 13.31$, $p = .001$).

Gruppenzentrierte Sympathieeinschätzung – Verkuyten-Sympathie-Item

Die Verkuyten-Sympathieeinschätzung wurde nur zum dritten Messzeitpunkt auch mit russischen Kindern als Zielfremdgruppe verwendet. Die Werte verteilen sich hier zwischen -1 bis 3 bei einem Wertebereich der Skala von -3 bis 3. Die Mittelwerte von Interventions- und Kontrollgruppe sind klein, aber deutlich verschieden von Null, was für eine leichte Eigengruppenbevorzugung gegenüber der Gruppe russischer Kinder spricht. Dabei zeigen die Teilnehmer der Kontrollgruppe eine etwas größere Einstellungsverzerrung gegenüber der Gruppe der russischen Kinder, als die Kinder der Interventionsgruppe. Dies zeigt sich auch an der kleinen negativen Effektstärke, wie in Tabelle 40 zu sehen ist.

Tabelle 40

Mittelwerte, Standardabweichungen und Netto-Effektstärken des Verkuyten-Sympathie-Items gegenüber der Gruppe russischer Kindern (r), getrennt nach Interventionsgruppe und Kontrollgruppe zum dritten Messzeitpunkt

		T3	
		M (SD)	d_{t3}
Verkuyten-Sympathie (r)	IG	1.02 (0.91)	-0.29*
	KG	1.31 (1.10)	

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß; * = signifikanter Mittelwertsunterschied mit $p < 0.05$.

Der Mittelwertunterschied zwischen Interventions- und Kontrollgruppe wird bei der Varianzanalyse auch inferenzstatistisch signifikant ($F [1, 400] = 7.89, p = .005$).

Soziale Distanz – Haus-Aufgabe

Die Werte der Eigengruppenbevorzugung in der sozialen Distanz gegenüber einer russischen Familie verteilen sich über die gesamte Bandbreite des möglichen Wertebereichs von -3 bis 3. Dabei indizieren positive Werte eine größere Distanz zur Fremdgruppe und negative Werte eine größere Distanz zur Eigengruppe.

Die Mittelwerte der Eingengruppenbevorzugung liegen insgesamt auf einem relativ niedrigen Niveau, sind aber positiv. Dabei liegen die Werte der

Interventionsgruppe unter denen der Kontrollgruppe. Die Mittelwerte bleiben zwischen t2 und t3 relativ stabil. Dies erklärt auch die etwa gleich großen negativen Effektstärken innerhalb des zweiten und innerhalb des dritten Messzeitpunktes. Mittelwerte, Standardabweichungen und Effektstärken finden sich in Tabelle 41.

Tabelle 41

Mittelwerte, Standardabweichungen und Effektstärken der Haus-Aufgabe gegenüber einer russischen (r) Familie, getrennt nach Interventionsgruppe und Kontrollgruppe zu den Messzeitpunkten t2 und t3

		T2	T3		
		<i>M</i>	<i>M</i>	<i>d_{t2}</i>	<i>d_{t3}</i>
		(<i>SD</i>)	(<i>SD</i>)		
Haus-Aufgabe (r)	IG	0.29 (1.34)	0.31 (1.19)	-0.43*	-0.38*
	KG	0.85 (1.26)	0.76 (1.21)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; *d* = Effektstärkemaß; * = signifikanter Mittelwertsunterschied mit $p < 0.05$.

Da die Haus-Aufgabe nicht zum Prätest durchgeführt wurde, finden in der Auswertung nur Vergleiche zwischen den Untersuchungsgruppen innerhalb eines Messzeitpunktes statt. Die oben dargestellten Mittelwertsunterschiede zwischen der Interventionsgruppe und der Kontrollgruppe erweisen sich in der durchgeführten multivariaten Varianzanalyse beide, sowohl zum zweiten ($F [1, 399] = 18.51$, $p = .000$), als auch zum dritten ($F [1, 399] = 14.16$, $p = .000$) Messzeitpunkt, als statistisch bedeutsam.

Chinesische Fremdgruppe

Personenzentrierte Sympathieeinschätzung – Intended Behavior Measure-Sympathie-Item

Die IBM-Sympathieeinschätzungen gegenüber einem chinesischen Kind im Vergleich zum deutschen Kind liegen in einem Bereich zwischen -4 und 4 und nutzen damit den Wertebereich der Skala vollständig aus. Die Mittelwerte in beiden Gruppen sind positiv und liegen nahe Null, was für eine minimale Eigengruppenbevorzugung der Teilnehmer spricht. Dabei unterscheiden sich die Prätestwerte der Untersuchungsgruppen nicht signifikant voneinander ($F [1, 394] = 2.22, p = .137$).

In beiden Untersuchungsgruppen findet ein leichter Anstieg der Eigengruppenbevorzugung über die Zeit statt. In der Interventionsgruppe liegt ein sehr leichter aber kontinuierlicher Anstieg vor, während die Kontrollgruppe vom ersten zum zweiten Messzeitpunkt stabil in ihren Werten bleibt und zum dritten Messzeitpunkt einen größeren Anstieg zu verzeichnen hat. Diese unterschiedlichen Entwicklungsverläufe machen sich an der kleinen positiven Netto-Effektstärke im Prä-Post-Vergleich und an der kleinen negativen Effektstärke im Prä-Follow-up-Vergleich bemerkbar. Mittelwerte, Standardabweichungen und Netto-Effektstärken lassen sich in Tabelle 42 ablesen.

Tabelle 42

Mittelwerte, Standardabweichungen und Netto-Effektstärken des IBM-Sympathie-Items gegenüber einem chinesischen Kind (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
IBM-Sympathie (c)	IG	0.13 (1.04)	0.30 (1.09)	0.41 (1.18)	0.18	-0.13
	KG	0.29 (1.09)	0.27 (1.04)	0.71 (1.03)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

Die Unterschiede im Entwicklungsverlauf werden auch in der inferenzstatistischen Analyse signifikant. So findet sich eine signifikante Interaktion der Faktoren Messzeitpunkte und Gruppe ($F [2, 392] = 3.25, p = .040$). Diese beruht nach Berechnung der Post-hoc-Kontraste auf unterschiedlichen Entwicklungen der beiden Untersuchungsgruppen zwischen dem zweiten und dritten Messzeitpunkt ($F [1, 393] = 4.04, p = .045$). Weder der Kontrast zwischen dem ersten und dem zweiten Messzeitpunkt ($F [1, 393] = 2.27, p = .133$), noch der Kontrast zwischen dem ersten und dem dritten Messzeitpunkt ($F [1, 393] = 0.91, p = .341$) weisen signifikante Unterschiede auf. Auch die Haupteffekte der Faktoren Messzeitpunkte ($F [2, 392] = 13.89, p = .000$) und Gruppe ($F [1, 393] = 4.10, p = .044$) werden statistisch bedeutsam.

Gruppenzentrierte Sympathieeinschätzung – Verkuyten-Sympathie-Item

Die Werte der Verkuyten-Sympathieeinschätzung gegenüber der Gruppe chinesischer Kinder verteilen sich auf einen Bereich von -2 bis 3 und nutzen damit den Wertebereich der Skala (-3 bis 3) fast komplett aus. Die Mittelwerte beider Untersuchungsgruppen liegen im kleineren positiven Bereich, was für eine geringe aber vorhandene Eigengruppenbevorzugung auf dieser Skala spricht.

Die Mittelwerte der beiden Untersuchungsgruppen im Prätest unterscheiden sich signifikant voneinander ($F [1, 398] = 5.41, p = .031$). Die Werte in der Interventionsgruppe liegen unterhalb der Werte in der Kontrollgruppe. Die Entwicklungsverläufe sind in beiden Untersuchungsgruppen aber relativ ähnlich. Zunächst findet ein Abfall der Einstellungsverzerrungen statt. Zwischen dem zweiten und dem dritten Messzeitpunkt gibt es dann wieder einen minimalen Anstieg der Eigengruppenbevorzugung. Die parallelen Entwicklungsverläufe spiegeln sich auch in den nahe Null liegenden Netto-Effektstärken wider. Mittelwerte, Standardabweichungen und Netto-Effektstärken lassen sich in Tabelle 43 ablesen.

Tabelle 43

Mittelwerte, Standardabweichungen und Netto-Effektstärken des Verkuyten-Sympathie-Items gegenüber der Gruppe chinesischer Kinder (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
Verkuyten-Sympathie (c)	IG	0.91	0.72	0.82	-0.01	0.05
		(1.07)	(1.01)	(0.92)		
	KG	1.15	0.97	1.01		
		(1.08)	(0.94)	(0.94)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

Da sich auf diesem Item signifikante Prätest-Unterschiede zwischen den beiden Untersuchungsgruppen finden, werden an dieser Stelle zwei Regressionen berechnet, die sowohl den Prätestwert, als auch die Interaktion des Prätests mit dem Faktor Gruppenzugehörigkeit als Regressoren neben dem Regressor Gruppe simultan berücksichtigen. Dies geschieht, um mögliche Verzerrungen durch die Prätest-Unterschiede bei der Beurteilung der Ergebnisse herauszurechnen.

In der Regression auf den Posttest des Verkuyten-Sympathie-Items gegenüber der Gruppe chinesischer Kinder erweisen sich in dem signifikanten Regressionsmodell ($F [3, 395] = 17.46, p = .000$) mit einem von $r^2 = 0.14$ sowohl der Regressor Gruppe ($\beta = .20, p = .003$), als auch der Regressor Prätest ($\beta = .64, p = .000$). neben dem Interaktionsterm der beiden Regressoren ($\beta = -.35, p = .026$) als statistisch bedeutsam.

In der zweiten Regression auf das Follow-up des Verkuyten-Sympathie-Items gegenüber der Gruppe chinesischer Kinder zeigen sich nur noch die Regressoren Prätest ($\beta = .61, p = .000$) und Gruppe ($\beta = .15, p = .026$) im signifikanten Regressionsmodell ($F [3, 395] = 16.91, p = .000$) mit einem $r^2 = 0.14$ signifikant. Der Interaktionsterm der beiden Regressoren wird jedoch nicht mehr signifikant ($\beta = -.29, p = .069$).

Soziale Distanz – Haus-Aufgabe

Die Werte der Eigengruppenbevorzugung in der sozialen Distanz gegenüber einer chinesischen Familie verteilen sich über die gesamte Bandbreite des möglichen Wertebereichs von -3 bis 3. Dabei indizieren positive Werte eine größere Distanz zur Fremdgruppe und negative Werte eine größere Distanz zur Eigengruppe.

Die Mittelwerte der Eigengruppenbevorzugung liegen insgesamt auf einem relativ niedrigen Niveau, sind aber positiv. Dabei liegen die Werte der Interventionsgruppe unter denen der Kontrollgruppe. Die Mittelwerte bleiben zwischen t2 und t3 relativ stabil. Dies erklärt auch die etwa gleich großen negativen Effektstärken innerhalb des zweiten und innerhalb des dritten Messzeitpunktes. Mittelwerte, Standardabweichungen und Effektstärken finden sich in Tabelle 44.

Tabelle 44

Mittelwerte, Standardabweichungen und Effektstärken der Haus-Aufgabe gegenüber einer chinesischen (c) Familie, getrennt nach Interventionsgruppe und Kontrollgruppe zu den Messzeitpunkten t2 und t3

		T2	T3	d_{t2}	d_{t3}
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
Haus-Aufgabe (c)	IG	0.43 (1.42)	0.34 (1.25)	-0.23*	-0.24*
	KG	0.74 (1.33)	0.63 (1.17)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß; * = signifikanter Mittelwertsunterschied mit $p < 0.05$.

Da die Haus-Aufgabe nicht zum Prätest durchgeführt wurde, finden in der Auswertung nur Vergleiche zwischen den Untersuchungsgruppen innerhalb eines Messzeitpunktes statt. Die oben dargestellten Mittelwertsunterschiede zwischen der Interventionsgruppe und der Kontrollgruppe erweisen sich in der durchgeführten multivariaten Varianzanalyse beide, sowohl zum zweiten ($F [1, 399] = 5.16, p = .024$), als auch zum dritten ($F [1, 399] = 5.86, p = .016$) Messzeitpunkt, als statistisch bedeutsam.

Zusammenfassung der Ergebnisse zu gruppenbezogenen sozialen

Einstellungen mit affektivem Schwerpunkt

Die gruppenbezogenen sozialen Einstellungen mit affektivem Schwerpunkt liegen, ebenso wie die gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt, insgesamt auf einem sehr niedrigen Level. Dabei fällt auch bei diesen Einstellungsmaßen ein leichter genereller Niveauunterschied zwischen der chinesischen und der russischen Fremdgruppe auf, wobei die Eigengruppenbevorzugung gegenüber der russischen Fremdgruppe stets etwas stärker ausgeprägt ist, als gegenüber der chinesischen Fremdgruppe. Nichts desto trotz liegen auch die Einstellungsverzerrungen gegenüber der russischen Gruppe immer noch auf einem sehr niedrigen Level. Insgesamt lässt sich in diesem Einstellungsbereich ein positives Bild der Interventionseffekte, vor allem gegenüber der russischen Fremdgruppe, zeichnen.

Auf dem IBM-Sympathie-Item gegenüber einem russischen Kind zeigt sich zum dritten Messzeitpunkt, ähnlich wie bei dem Verkuyten-Sympathie-Item (s.u.), ein signifikanter Niveauunterschied zwischen den Untersuchungsgruppen. Die Teilnehmer der Interventionsgruppe zeigen hier weniger Einstellungsverzerrungen gegenüber der russischen Fremdgruppe als die Teilnehmer der Kontrollgruppe. Im Mittelwertvergleich der Einstellungsverzerrungen gegenüber der Gruppe russischer Kinder auf den Verkuyten-Sympathie-Einschätzungen findet sich ebenfalls ein signifikanter Niveauunterschied zum dritten Messzeitpunkt. Die Interventionsgruppe zeigt auch hier weniger Eigengruppenbevorzugung gegenüber der Gruppe russischer Kinder.

Die Mittelwertunterschiede in der sozialen Distanz zeigen ebenfalls, dass die Interventionsgruppe weniger affektive Eigengruppenbevorzugung gegenüber der russischen Zielfremdgruppe inne hat. So finden sich auch hier signifikante Mittelwertunterschiede zwischen den Untersuchungsgruppen gegenüber einer russischen Familie. Diese Unterschiede bestehen innerhalb des zweiten und des dritten Messzeitpunktes.

Das IBM-Sympathie-Item gegenüber einem chinesischen Kind zeigt im Gegensatz zum Verkuyten-Sympathie-Item (s.u.) einen leichten Anstieg der Werte über die Zeit in beiden Gruppen. Hier nimmt zunächst die Interventionsgruppe stärker zu als die Kontrollgruppe, was sich zum Follow-up hin allerdings umkehrt. Da bleibt die Interventionsgruppe in ihren Werten stabil, während die Kontrollgruppe zunimmt. Die

Interventionsgruppe zeigt also verzögert einen geringeren Anstieg als die Kontrollgruppe. Diese unterschiedlichen Entwicklungsverläufe werden auch statistisch signifikant. Eine ähnliche Entwicklung zeichnet sich auch in der entsprechenden Risikogruppe 25 ab, ohne dabei allerdings eine signifikante Interaktion zu erzeugen (s. Anhang D).

Bei dem Verkuyten-Sympathie-Item gegenüber der Gruppe chinesischer Kinder finden sich hingegen keine bedeutsamen Unterschiede in der Entwicklung der beiden Untersuchungsgruppen. Dies gilt auch für die entsprechende Risikogruppe 25 (s. Anhang D).

Die Mittelwertunterschiede in der sozialen Distanz zeigen jedoch, dass die Interventionsgruppe auch weniger affektive Eigengruppenbevorzugung gegenüber der chinesischen Fremdgruppe innehat. So zeigen sich hier signifikante Gruppenunterschiede gegenüber einer chinesischen Familie. Diese Unterschiede finden sich sowohl innerhalb des zweiten als auch des dritten Messzeitpunktes, sind aber etwas kleiner als die Gruppenunterschiede in der sozialen Distanz gegenüber der russischen Zielfremdgruppe.

Insgesamt kann festgehalten werden, dass es signifikante Unterschiede zwischen den Untersuchungsgruppen im Hinblick auf die gruppenbezogenen sozialen Einstellungen mit affektivem Schwerpunkt gibt. Diese zeigen sich allerdings hauptsächlich gegenüber der russischen Zielfremdgruppe. Die Effekte sind dabei eher klein, aber statistisch bedeutsam, weisen in die erwartete Richtung und sind konsistent über verschiedene Maße hinweg auffindbar.

IV.4.5. Gruppenbezogene soziale Einstellungen mit konativem

Schwerpunkt

Russische Zielfremdgruppe

Kontaktintentionen – Intended Behavior Measure

Die Werte der Eigengruppenbevorzugung der Kontaktintentionen gegenüber einem russischen Kind wurden nur zum dritten Messzeitpunkt erhoben. Sie verteilen sich über den gesamten Wertebereich der Skala von -4 bis 4. Die Mittelwerte der beiden Untersuchungsgruppen sind relativ klein aber positiv. Dabei ist der Wert in der Kontrollgruppe etwas höher als der Wert in der Interventionsgruppe. Dies lässt sich auch an der kleinen negativen Effektstärke erkennen. Effektstärken, Mittelwerte und Standardabweichungen sind in Tabelle 45 abzulesen.

Tabelle 45

Mittelwerte, Standardabweichungen und Netto-Effektstärken der IBM-Kontaktintention-Eigengruppenbevorzugung (IBM-K-EGB) gegenüber einem russischen Kind (r), getrennt nach Interventionsgruppe und Kontrollgruppe zum dritten Messzeitpunkt

		T3	
		<i>M (SD)</i>	<i>d_{t3}</i>
IBM-K-EGB (r)	IG	0.33 (1.08)	-0.35*
	KG	0.72 (1.14)	

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; *d* = Effektstärkemaß; * = signifikanter Mittelwertsunterschied mit $p < 0.05$.

Dieser Mittelwertsunterschied zwischen den beiden Untersuchungsgruppen wird auch statistisch bedeutsam ($F [1, 400] = 12.77, p = .000$).

Chinesische Fremdgruppe

Kontaktintentionen – Intended Behavior Measure

Die Werte der Kontaktintention gegenüber einem chinesischen Kind verteilen sich über den kompletten Wertebereich der Skala von -4 bis 4. Die kleinen positiven Mittelwerte in beiden Gruppen bilden eine leichte Eigengruppenbevorzugung in den

Kontaktintentionen der Teilnehmer ab. Die Prätestmittelwerte der beiden Untersuchungsgruppen unterscheiden sich nicht signifikant ($F [1, 398] = 2.42$, $p = .120$), auch wenn das Niveau der Werte in der Interventionsgruppe leicht niedriger ist als in der Kontrollgruppe.

Bei beiden Untersuchungsgruppen lässt sich ein leichter Anstieg in der Eigengruppenbevorzugung bei der Kontaktintention über die Zeit feststellen. Dabei steigt der Mittelwert in der Interventionsgruppe von t1 zu t2 zunächst auf das Niveau der Kontrollgruppe an, während die Kontrollgruppe relativ stabil bleibt. Vom zweiten zum dritten Messzeitpunkt steigt der Mittelwert in der Interventionsgruppe weiter leicht an, während es in der Kontrollgruppe einen deutlich stärkeren Anstieg gibt. Diese Entwicklungsverläufe führen auch zu den kleinen erst positiven (Prä-Post-Vergleich), dann negativen (Prä-Follow-up-Vergleich) Netto-Effektstärken. Mittelwerte, Standardabweichungen und Netto-Effektstärken sind in Tabelle 46 abzulesen.

Tabelle 46

Mittelwerte, Standardabweichungen und Netto-Effektstärken der IBM-Kontaktintention-Eigengruppenbevorzugung (IBM-K_EGB) gegenüber einem chinesischen Kind (c), getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		M (SD)	M (SD)	M (SD)		
IBM-K-EGB (c)	IG	0.10	0.28	0.42	0.13	-0.19
		(1.00)	(1.14)	(1.14)		
	KG	0.27	0.32	0.79		
		(1.08)	(1.10)	(1.16)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

Inferenzstatistisch wird die Interaktion der Faktoren Messzeitpunkte und Gruppe nicht signifikant ($F [2, 397] = 2.92$, $p = .055$). Die beiden Haupteffekte der einzelnen Faktoren werden auch statistisch bedeutsam. Dabei spiegelt der Haupteffekt des Faktors Messzeitpunkte den Anstieg beider Untersuchungsgruppen in der Eigengruppenbevorzugung wider ($F [2, 397] = 20.83$, $p = .000$), während der

Haupteffekt des Faktors Gruppe ($F [1, 398] = 6.30, p = .013$) die übergreifenden Niveau-Unterschiede zwischen den Untersuchungsgruppen bestätigt.

Zusammenfassung der Ergebnisse zu gruppenbezogenen sozialen

Einstellungen mit konativem Schwerpunkt

Die Werte der gruppenbezogenen sozialen Einstellungen mit konativem Schwerpunkt bewegen sich ähnlich wie in den anderen beiden gruppenbezogenen Einstellungsbereichen insgesamt auf einem sehr niedrigen Niveau.

Die gruppenbezogenen sozialen Einstellungen mit konativem Schwerpunkt gegenüber einem russischen Kind zeigen die auch in anderen Einstellungsbereichen gefundenen Mittelwertsunterschiede zwischen der Interventionsgruppe und der Kontrollgruppe. Die Interventionsgruppe zeigt hier signifikant weniger Eigengruppenbevorzugung gegenüber dem russischen Kind als die Kontrollgruppe. Die Effektstärke ist zwar klein, aber vorhanden.

Im Vergleich dazu scheint bei den konativen Einstellungsverzerrungen gegenüber der chinesischen Fremdgruppe über die Zeit ein leichter Anstieg statt zu finden. Dies zeigt sich sowohl in der Gesamtgruppe als auch in der Risikogruppe 25, auch wenn hier zwischen Prä- und Posttest ein deutlicher Abfall der Werte zu verzeichnen ist. Zwischen Posttest und Follow-up nehmen auch hier die Werte wieder zu. Es finden sich jedoch keine statistisch bedeutsamen Interaktionen der Faktoren Messzeitpunkte und Gruppe (s. Anhang D). Allerdings zeigt die Interventionsgruppe zwischen Prä- und Posttest im Vergleich zur Kontrollgruppe etwas mehr Einstellungsverzerrungen. Dies ändert sich jedoch zum Follow-up hin wieder.

Insgesamt zeigt sich ein kleiner aber signifikanter Effekt auf der konativen Einstellungsskala gegenüber Kindern der russischen Fremdgruppe. Es finden sich jedoch keine signifikanten Effekte gegenüber Kindern der chinesischen Fremdgruppe.

IV.4.6. Interpersonale soziale Einstellungen

Toleranz – Fragebogen zur interpersonalen Toleranz

Die Werte der beiden Subskalen zur interpersonalen Toleranz verteilen sich unterschiedlich über den Wertebereich von 1 bis 5. Die Werte der Skala Intoleranz nutzt den gesamten Wertebereich aus, während die Werte der Subskala Akzeptanz sich über einen Bereich von 2.33 bis 5 erstrecken. Die unteren Bereiche dieser Subskala werden von den Teilnehmern nicht genutzt. Im Folgenden werden die Ergebnisse der beiden Subskalen nacheinander beschrieben.

Akzeptanz. Die Mittelwerte der Subskala Akzeptanz liegen bereits zum ersten Messzeitpunkt sehr hoch. Dabei unterscheiden sich die Mittel der beiden Untersuchungsgruppen nicht signifikant voneinander ($F [1, 398] = 0.12, p = .726$). Zum zweiten und dritten Messzeitpunkt ist ein sehr leichter Anstieg in beiden Untersuchungsgruppen festzustellen. Dieser ist in der Interventionsgruppe minimal größer als in der Kontrollgruppe, was die sehr kleinen positiven Effektstärken erklärt.

Intoleranz. Die Mittelwerte der Subskala Intoleranz liegen eher in der Mitte des Wertebereichs und unterscheiden sich zum ersten Messzeitpunkt nicht zwischen den Untersuchungsgruppen ($F [1, 398] = 0.46, p = .498$). Über die Zeit sinkt die Intoleranz in beiden Untersuchungsgruppen ab. Wobei in der Interventionsgruppe zunächst ein stärkerer Abfall zu verzeichnen ist, was die kleine negative Prä-Post-Netto-Effektstärke bedingt. Dieser Unterschied nivelliert sich jedoch zum dritten Messzeitpunkt, in dem sich beide Untersuchungsgruppen wieder auf einem ähnlichen Niveau einpegeln, was auch an der reduzierten Netto-Effektstärke sichtbar wird. Mittelwerte, Standardabweichungen und Netto-Effektstärken finden sich in Tabelle 47.

Tabelle 47

Mittelwerte, Standardabweichungen und Netto-Effektstärken des Fragebogens zur interpersonalen Toleranz, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
Toleranz (A)	IG	4.40 (0.46)	4.50 (0.39)	4.49 (0.41)	0.10	0.06
	KG	4.42 (0.52)	4.47 (0.45)	4.48 (0.41)		
Toleranz (I)	IG	2.87 (0.88)	2.26 (0.80)	2.15 (0.77)	-0.36*	-0.09
	KG	2.81 (0.86)	2.51 (0.83)	2.17 (0.74)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; A = Subskala Akzeptanz; I = Subskala Intoleranz; d = Effektstärkemaß; * = signifikant unterschiedlicher Entwicklungsverlauf mit $p < 0.05$.

Akzeptanz. Inferenzstatistisch bestätigt sich der deskriptive Eindruck. Es findet sich keine signifikante Interaktion der Faktoren Messzeitpunkte und Gruppe ($F [2, 397] = 0.31, p = .735$). Auch der Haupteffekt des Faktors Gruppe wird nicht statistisch bedeutsam ($F [1, 398] = 0.07, p = .794$). Lediglich der Haupteffekt des Faktors Messzeitpunkte wird signifikant ($F [2, 397] = 4.13, p = .017$).

Intoleranz. Bei der Subskala Intoleranz findet sich hingegen eine signifikante Interaktion der Faktoren Messzeitpunkte und Gruppe ($F [2, 397] = 8.91, p = .000$). Dabei wird der Post-hoc-Kontrast zwischen dem ersten und dem zweiten Messzeitpunkt statistisch bedeutsam ($F [1, 398] = 12.56, p = .000$). Die Kontraste zwischen dem ersten und dritten Messzeitpunkt ($F [1, 398] = 0.57, p = .449$) und zwischen dem zweiten und dritten Messzeitpunkt ($F [1, 398] = 1.52, p = .219$) erreichen nicht das Signifikanzniveau. Zusätzlich wird auch hier der Haupteffekt des Faktors Messzeitpunkte signifikant ($F [2, 397] = 115.45, p = .000$). Der Haupteffekt des Faktors Gruppe wird nicht statistisch bedeutsam ($F [1, 398] = 0.97, p = .326$).

Sozialverhalten

Verhaltenseinschätzungen durch Eltern –

Strengths and Difficulties Questionnaire

Gesamtproblemwert. Die Gesamtproblemwerte der Elterneinschätzungen des Sozialverhaltens ihrer Kinder reichen von 0 bis 20. Damit nutzen die Eltern nur die unteren beiden Drittel des gesamten Wertebereichs der Skala aus (0-30).

Die Mittelwerte des Prätests unterscheiden sich nicht statistisch bedeutsam zwischen den beiden Untersuchungsgruppen ($F [1, 380] = 0.58, p = .448$). Die Interventionsgruppe nimmt über die Zeit leicht im Gesamtproblemwert ab, während die Kontrollgruppe hier relativ stabil bleibt. So entstehen auch die sehr kleinen negativen Netto-Effektstärken. Insgesamt bleiben die Werte über die Zeit recht stabil.

Prosoziales Verhalten. Die Werte der Elterneinschätzungen im prosozialem Verhalten ihres Kindes verteilen sich von 2.5 bis 10 bei einem Wertebereich von 0 bis 10. Die Mittelwerte der beiden Untersuchungsgruppen unterscheiden sich zum Prätest nicht signifikant ($F [1, 380] = 0.55, p = .460$). In der Interventionsgruppe gibt es von t1 zu t2 einen minimalen Anstieg. Im Anschluss daran bleibt der Wert bei dieser Gruppe in etwa stabil. Bei der Kontrollgruppe erfolgt ebenfalls ein Anstieg zwischen Prä- und Posttest. Dieser setzt sich in abgeschwächter Form aber auch zum Follow-up hin fort. Der Anstieg in der Kontrollgruppe ist leicht größer als in der Interventionsgruppe. Daher rühren auch die sehr kleinen negativen Netto-Effektstärken.

Mittelwerte, Standardabweichungen und Netto-Effektstärken beider Subskalen finden sich in Tabelle 48.

Tabelle 48

Mittelwerte, Standardabweichungen und Netto-Effektstärken des SDQ, eingeschätzt durch die Eltern, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	T3	d_{t2-t1}	d_{t3-t1}
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
SDQ Gesamtproblemwert	IG	4.78 (3.56)	4.69 (3.57)	4.40 (3.36)	-0.02	-0.11
	KG	4.52 (3.29)	4.48 (3.39)	4.53 (3.58)		
SDQ prosoziales Verhalten	IG	8.22 (1.35)	8.36 (1.35)	8.33 (1.33)	-0.06	-0.12
	KG	8.12 (1.50)	8.35 (1.44)	8.40 (1.32)		

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß.

Gesamtproblemwert. Hier finden sich weder eine signifikante Interaktion der Faktoren Messzeitpunkte und Gruppe ($F [2, 351] = 1.40, p = .248$), noch signifikante Haupteffekte des Faktors Messzeitpunkte ($F [2, 351] = 0.38, p = .687$) oder Gruppe ($F [1, 352] = 0.03, p = .856$).

Prosoziales Verhalten. Es findet sich keine statistisch bedeutsame Interaktion der Faktoren Messzeitpunkte und Gruppe ($F [2, 351] = 0.50, p = .610$). Ebenso wenig findet sich ein bedeutsamer Haupteffekt des Faktors Gruppe ($F [1, 352] = .03, p = .873$). Allein der Haupteffekt des Faktors Messzeitpunkte wird signifikant und bestätigt damit den leichten Anstieg im prosozialem Verhalten in beiden Untersuchungsgruppen ($F [2, 351] = 6.42, p = .002$).

Sozialverhalten

Verhaltenseinschätzungen durch Lehrer – Strengths and Difficulties Questionnaire

Um die positive Zusammenarbeit mit den Lehrern der Schulen nicht zu belasten, wurden diese nur zum Prätest und zum Posttest, aber nicht zum Follow-up befragt.

Gesamtproblemwert. Die Werte der Lehrereinschätzungen des Gesamtproblemwerts verteilen sich von 0 bis 27, womit fast die gesamte Breite der zur Verfügung stehenden Skala von 0 bis 30 ausgeschöpft wird. Die Mittelwerte liegen jedoch in beiden Untersuchungsgruppen auf einem sehr geringen Wert. Dabei unterscheiden sich die Untersuchungsgruppen im Prätest nicht signifikant voneinander ($F [1, 397] = 0.50, p = .478$). Zum Posttest hin findet ein ganz leichter Abfall des Gesamtproblemwerts statt, wobei die Kontrollgruppe etwas stärker in ihren Werten absinkt als die Interventionsgruppe. Dies führt zu der minimalen positiven Netto-Effektstärke.

Prosoziales Verhalten. Die Werte der Lehrereinschätzungen des prosozialen Verhaltens der Kinder verteilen sich zwischen 1 und 10 und somit fast über den gesamten Wertebereich von 0 bis 10. Die Mittelwerte liegen in beiden Untersuchungsgruppen auf relativ hohen Werten. Dabei unterscheiden sich die Prätestwerte der Gruppen nicht signifikant voneinander ($F [1, 397] = 0.14, p = .706$).

Die Entwicklungsverläufe der Untersuchungsgruppen zwischen Prä- und Posttest sind jedoch unterschiedlich. In der Interventionsgruppe findet ein leichter Anstieg im prosozialen Verhalten statt, während in der Kontrollgruppe ein leichter Abfall der Werte zu beobachten ist. Dies führt zu der kleinen positiven Netto-Effektstärke.

Mittelwerte, Standardabweichungen und Netto-Effektstärken finden sich in Tabelle 49.

Tabelle 49

Mittelwerte, Standardabweichungen und Netto-Effektstärken des SDQ, eingeschätzt durch die Lehrer, getrennt nach Interventionsgruppe und Kontrollgruppe über die Messzeitpunkte hinweg

		T1	T2	d_{t2-t1}
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	
SDQ Gesamtproblem	IG	4.03 (4.18)	3.95 (3.87)	0.03
	KG	3.74 (3.82)	3.54 (3.96)	
SDQ prosoziales Verhalten	IG	7.78 (2.12)	8.13 (1.93)	0.33*
	KG	7.86 (2.10)	7.52 (2.37)	

Anmerkungen. IG = Interventionsgruppe; KG = Kontrollgruppe; d = Effektstärkemaß; * = signifikant unterschiedlicher Entwicklungsverlauf mit $p < 0.05$.

Gesamtproblemwert. Inferenzstatistisch wird hier weder die Interaktion der Faktoren Messzeitpunkte und Gruppe ($F [1, 397] = 0.14, p = .710$), noch der Haupteffekt des Faktors Gruppe ($F [1, 397] = 0.88, p = .350$), oder der Haupteffekt des Faktors Messzeitpunkte ($F [1, 397] = 0.55, p = .457$) bedeutsam. Die Werte bleiben weitestgehend stabil.

Prosoziales Verhalten. Beim prosozialem Verhalten hingegen bestätigen sich die unterschiedlichen Entwicklungsverläufe mit einer signifikanten Interaktion der Faktoren Messzeitpunkte und Gruppe ($F [1, 397] = 13.23, p = .000$). Ansonsten werden weder der Haupteffekt des Faktors Gruppe ($F [1, 397] = 1.79, p = .182$), noch der Haupteffekt des Faktors Messzeitpunkte statistisch bedeutsam ($F [1, 397] = 0.02, p = .887$).

Zusammenfassung der Ergebnisse zu interpersonalen sozialen Einstellungen und Sozialverhalten

Insgesamt zeigt sich in den interpersonalen sozialen Einstellungen und dem generellen Sozialverhalten der Kinder ein hohes Niveau bei positiven Einstellungen und prosozialen Verhaltenseinschätzungen sowie ein niedriges Niveau bei den negativen Einstellungen und den problematischen Verhaltenseinschätzungen. Die Werte bei den interpersonalen Einstellungen beispielsweise liegen im Bereich Akzeptanz auf einem sehr hohen Niveau. Es findet sich zudem ein minimaler Anstieg über die Zeit in beiden Untersuchungsgruppen, aber keine bedeutsamen Interventionseffekte, was einem möglichen Deckeneffekt der Skala geschuldet sein könnte. Im Bereich Intoleranz hingegen zeigt sich trotz des gemeinsamen Abfalls der Intoleranzwerte ein kurzfristiger signifikanter Interventionseffekt. Hier sinken die Werte der Interventionsgruppe zwischen t1 und t2 stärker ab als in der Kontrollgruppe. Dieser Unterschied nivelliert sich jedoch zum Follow-up wieder. Ein ähnlicher Beschleunigungseffekt findet sich auch bei der Analyse der Risikogruppe 25 im Bereich der Akzeptanz. Hier steigt der Wert in der Interventionsgruppe zunächst stärker an als in der Kontrollgruppe. Zum Follow-up hin gleichen sich die Werte jedoch wieder an. Allerdings wird der Effekt hier statistisch nicht bedeutsam. Genauso scheint es im Bereich der Intoleranz bei der entsprechenden Risikogruppe 25 zu sein. Auch hier findet sich ein nicht-signifikanter Beschleunigungseffekt in der Interventionsgruppe (s. Anhang D).

Beim Sozialverhalten – eingeschätzt durch die Eltern der Teilnehmer – findet sich im Bereich des Gesamtproblemwerts in beiden Untersuchungsgruppen ein Abfall der Werte über die Zeit. Dieser ist in der Interventionsgruppe leicht stärker als in der Kontrollgruppe. Der Unterschied in den Entwicklungsverläufen erreicht jedoch nicht das Signifikanzniveau. Beim Bereich des prosozialen Verhaltens findet sich ein Anstieg der Werte über die Zeit. Dabei steigen die Werte in der Interventionsgruppe zunächst stärker an, als in der Kontrollgruppe. Dieser Unterschied gleicht sich jedoch zum Follow-up hin wieder aus. Auch hier findet sich also ein Beschleunigungseffekt der Intervention. In den entsprechenden Risikogruppen 25 beider Bereiche vollziehen sich äquivalente Entwicklungen wie in der oben beschriebenen Gesamtgruppe (s. Anhang D).

Bei den Verhaltenseinschätzungen der Lehrer der Teilnehmer zeigt sich beim Gesamtproblemwert ein minimaler Abfall in beiden Untersuchungsgruppen vom

ersten zum zweiten Messzeitpunkt. Im prosozialen Verhalten hingegen findet sich ein leichter Anstieg der Werte der Interventionsgruppe, während die Werte der Kontrollgruppe leicht absinken. Diese signifikant unterschiedlichen Entwicklungsverläufe der beiden Untersuchungsgruppen führen zu einer kleinen positiven Netto-Effektstärke im Prä-Post-Vergleich. Eine ähnliche Entwicklung findet sich in den Risikogruppen 25. Allerdings zeigt sich hier keine signifikante Interaktion im Bereich des prosozialen Verhaltens, da sich hier die Interventionsgruppe und die Kontrollgruppe verbessern, auch wenn die Interventionsgruppe sich mehr steigert als die Kontrollgruppe (s. Anhang).

Insgesamt zeigen sich bei den interpersonalen sozialen Einstellungen und beim Sozialverhalten kurzfristig positive Effekte der Intervention. Die Intoleranz nimmt stärker ab als in der Kontrollgruppe, das prosoziale Verhalten steigt – zumindest nach der Lehrereinschätzung – stärker an, als in der Kontrollgruppe. Hier scheinen Beschleunigungseffekte der Intervention vorzuliegen.

IV.4.7. Wirksamkeitsanalyse – Zusammenfassung

Um die Ergebnisse der Wirksamkeitsanalyse des Präventionsansatzes auf einen Blick erfassen zu können, werden diese unten in den Tabellen 50 und 51 noch einmal kurz und knapp zusammengefasst. Dabei werden zum einen die trendmäßigen Entwicklungen dargestellt, indem Richtung und Größe der Effektstärken gezeigt werden. Zum anderen wird die statistische Signifikanz der Ergebnisse dokumentiert. Die Tabelle ist parallel zur Hypothesentabelle im Kapitel III.8. Hypothesen der PARTS Studie aufgebaut. Somit ist ein Abgleich über den Erfolg von PARTS leicht möglich.

Tabelle 50

Zusammenfassung kurzfristiger und langfristiger Interventionseffekte der proximalen Maße

Konstrukt	Skala	Kurzfristige Effekte	Langfristige Effekte	Sig.
Interkulturelles Wissen	Wissenstest	0.78*	1.14*	√
Multiple Klassifikationsfähigkeit	Klassifikations-test	0.13	0.16	-
Perspektivenübernahmefähigkeit	Chandler Cartoons	0.18	0.20	-
Soziale Informationsverarbeitung	Test zum sozialen Problemlösen	-0.04	-0.03	-

Anmerkungen. Kurzfristige Effekte = d_{t2-t1} ; Langfristige Effekte = d_{t3-t1} ; Sig. = signifikante Interaktion der Faktoren Messzeitpunkte und Gruppe mit $p < 0.05$; * = Signifikanter Kontrast mit $p < 0.05$.

Tabelle 51

Zusammenfassung kurzfristiger und langfristiger Interventionseffekte der distalen Maße

Konstrukt	Skala	Kurzfristige Effekte	Langfristige Effekte	Sig.
Gruppenbezogene soziale Einstellungen (kognitiv)	MRA (EGB Russen)	0.19	-0.13	√
	MRA (FGA Russen)	0.08	-0.21	√
	Verkuyten-Skala (Russen)	-	-0.29*#	-
	MRA (EGB Chinesen)	0.18	0.13	-
	MRA (FGA Chinesen)	0.09	0.01	-
	Verkuyten-Skala (Chinesen)	0.02	0.12	-
Gruppenbezogene soziale Einstellungen (affektiv)	IBM-Symp. (Russen)		-0.32*#	-
	Verkuyten-Symp. (Russen)		-0.29*#	-
	Haus-Aufgabe (Russen)	-0.43*#	-0.38*#	-
	IBM-Symp. (Chinesen)	0.18	-0.13	√
	Verkuyten-Symp. (Chinesen)	-0.01	0.05	-

Konstrukt	Skala	Kurzfristige Effekte	Langfristige Effekte	Sig.
Gruppenbezogene soziale Einstellungen (affektiv)	Haus-Aufgabe (Chinesen)	-0.23*#	-0.24*#	-
Gruppenbezogene soziale Einstellungen (konativ)	IBM (Russen)		-0.35*#	-
	IBM (Chinesen)	0.13	-0.19	-
Interpersonale Toleranz	Toleranzskala (Intoleranz)	-0.36*	-0.09	√
	Toleranzskala (Akzeptanz)	0.10	0.06	-
Sozialverhalten (Eltern)	SDQ soziale Kompetenz	-0.06	-0.12	-
	SDQ Problemwert	-0.02	-0.11	-
Sozialverhalten (Lehrer)	SDQ soziale Kompetenz	0.33*	-	√
	SDQ Problemwert	0.03	-	-

Anmerkungen. Kurzfristige Effekte = d_{t2-t1} ; Langfristige Effekte = d_{t3-t1} ; Sig. = signifikante Interaktion der Faktoren Messzeitpunkte und Gruppe mit $p < 0.05$; * = Signifikanter Kontrast mit $p < 0.05$;

= Effektstärke innerhalb eines Messzeitpunktes; EGB = Eigengruppenbevorzugung;

FGA = Fremdgruppenabwertung.

V. Diskussion

Im Folgenden Abschnitt werden zunächst die Ziele der vorliegenden Arbeit nochmals aufgegriffen. Danach werden die Ergebnisse aus der Implementations- und der Wirksamkeitsanalyse zusammengefasst, auf Konstruktebene beschrieben und im Anschluss in den Kontext der aktuellen Forschungsliteratur eingebettet und diskutiert. Dazu werden mögliche Erklärungsansätze für das vorgefundene Ergebnismuster aufgezeigt. Die Ergebnisse werden mit den Zielen des PARTS-Programms abgeglichen, bevor schließlich auf Grenzen und Möglichkeiten der dieser Arbeit zu Grunde liegenden Studie eingegangen wird und die Implikationen für die Forschung, aber auch für die Praxis der Vorurteilsprävention, zusammenfassend erörtert werden.

V.1. *Ziele der vorliegenden Arbeit*

Ziel der vorliegenden Arbeit ist es, das multimodale Präventionsprogramm PARTS zu evaluieren und Aussagen über seine Umsetzbarkeit und seine Wirksamkeit treffen zu können. Zu diesem Zweck wurde die oben beschriebene aufwendige Evaluationsstudie entwickelt und durchgeführt.

Aussagen über die *Umsetzbarkeit von PARTS* lassen sich aus den Ergebnissen der Implementationsanalyse ableiten. Dabei sollen neben einer allgemeinen Einordnung der strukturellen Merkmale des Trainingsprogramms in den Forschungskontext auch spezifischere Fragestellungen, durch die Einschätzung der Umsetzbarkeit anhand der erhobenen Implementationsparameter beantwortet werden: Welche Programmteile ließen sich besonders gut umsetzen, bei welchen gab es Schwierigkeiten? Ist das Programm von allen Kindern gut aufgenommen worden, oder gab es auch Problemfälle? Bei welchen Teilnehmern konnte das Programm besonders gut durchgeführt werden, bei welchen war es schwieriger?

Aussagen über die *Wirksamkeit von PARTS* lassen sich aus den Ergebnissen der Wirksamkeitsanalyse ablesen. Hier geht es um den Erfolg des Programms in der Erreichung der zuvor festgesetzten Zielkriterien. Dazu wird zunächst geprüft, ob die proximalen Ziele von PARTS erreicht wurden. Wurden die individuellen sozial-

kognitiven Kompetenzen durch das Programm gefördert? Wurde das interkulturelle Wissen gefördert? Konnte der stellvertretende Kontakt gut umgesetzt werden?

Die distalen Ziele von PARTS bestanden in der positiven Beeinflussung von gruppenbezogenen sozialen Einstellungen mit kognitivem, affektivem und konativem Schwerpunkt. Weiterhin sollte die interpersonale Toleranz und das allgemeine Sozialverhalten durch das Programm positiv beeinflusst werden.

Dabei sollten die proximalen und distalen Ziele bei möglichst allen Teilnehmern erreicht werden, unabhängig von Alter, Geschlecht, Entwicklungskontext und Intelligenz (Unabhängigkeit der Effekte). Der positive Effekt sollte sich auch bei Kindern aus Risikogruppen bemerkbar machen (Risikogruppeneffekte) und sowohl kurzfristig als auch langfristig messbar sein (zeitliche Persistenz der Effekte). Schließlich sollte sich der Effekt auch auf andere Zielfremdgruppen generalisieren (Generalisierung der Effekte).

Die Beantwortung dieser Fragen zur Umsetzbarkeit und Wirksamkeit des Präventionsprogramms PARTS wird in den folgenden Passagen vorgenommen.

Doch zuvor wird geprüft, ob die Anforderungen an ein wissenschaftlich fundiertes Interventionsprogramm zur Prävention von Vorurteilen und Diskriminierung, die aus den Erkenntnissen bisheriger Forschung abgeleitet wurden (s. Kapitel II.4. Schlussfolgerungen für PARTS), bei der Entwicklung und Umsetzung von PARTS eingehalten wurden.

V.2. Anforderungen an ein wissenschaftlich fundiertes Programm

Präventives Vorgehen

Den abgeleiteten Anforderungen entsprechend verfolgt PARTS einen präventiven Ansatz und setzt in einer Phase erhöhter Plastizität der Einstellungsentwicklung ein. So wurde das Programm mit Kindern der dritten Klasse durchgeführt, deren Einstellungen sich noch nicht verfestigt haben (vgl. Raabe & Beelmann, in Druck), und die in der mittleren Kindheit, angetrieben von der kognitiven Entwicklung, auch normativ Veränderungen in ihren gruppenbezogenen sozialen Einstellungen erfahren (Aboud, 1988).

Ganzheitlicher (multimodaler) Ansatz

Um wie gefordert, möglichst viele Facetten von gruppenbezogenen sozialen Einstellungen zu beeinflussen, verfolgt PARTS einen multimodalen ganzheitlichen Interventionsansatz und verfügt über unterschiedliche Interventionsmethoden (Förderung sozial-kognitiver Kompetenzen, Wissensbasierte Intervention und Stellvertretender-Kontakt-Intervention), die geeignet sind, unterschiedliche Schwerpunkte gruppenbezogener sozialer Einstellungen zu erreichen (Dovidio et al., 2004). Dabei verfügt PARTS über den für die mittlere Kindheit geforderten sozial-kognitiven Fokus (Teichman & Bar-Tal, 2008).

Intensität und Dauer des Programms

Die geforderte Dauer des Trainingsprogramms, um zumindest kleine, möglichst längerfristige Effekte auf gruppenbezogenen sozialen Einstellungen zu zeigen, liegt bei mindestens 8 Sitzungen (Beelmann & Heinemann, in Vorbereitung, s. Kapitel II.4. Schlussfolgerungen für PARTS). PARTS umfasst mit 14 Sitzungen nahezu doppelt so viele Interventionseinheiten, als diese Forderung vorgibt (s. Kapitel III.2. Ablauf und Durchführung der Intervention). Somit ist der Anspruch dieser Untergrenze für wirksame Interventionen mehr als erfüllt.

Standardisierung und Strukturierung

Die geforderte Standardisierung und das hochstrukturierte Vorgehen zur besseren Evaluation und Wirksamkeit des Trainings wurde bei PARTS durch die Verwendung eines standardisierten Interventionsmanuals, standardisierter

Materialien, durch die Implementationskontrollen sowie durch internen Austausch und Supervisionssitzungen der Trainer untereinander sicher gestellt (s. Kapitel III. Methoden).

Bebildertes entwicklungsangemessenes Interventionsmaterial

Um eine möglichst wirksame und entwicklungsangemessene Intervention mit Drittklässlern durchführen zu können, verfügt PARTS über sehr stark bebildertes Interventionsmaterial. Es wurden bei fast allen Übungen und Spielen Bilder verwendet (s. Anhang A). Weiterhin wurden Geschichten und Materialien eingesetzt die zum Großteil bereits zuvor auf ihre Altersangemessenheit getestet wurden (s. Kapitel III. Methoden).

Kindzentriertheit

Diese Forderung an ein wissenschaftlich fundiertes Programm konnte bei PARTS nur bedingt erfüllt werden. So wurden die Kinder während der Interventionssitzungen immer wieder nach ihren eigenen Erfahrungen und Meinungen zu einzelnen Themen befragt (s. Anhang A). Allerdings konnte dafür in den einzelnen Interventionseinheiten nur ein begrenzter Zeitrahmen eingeräumt werden, um die manualgetreue Umsetzung des Programms auf Grund des engen Zeitrahmens nicht zu gefährden.

Sozialer Kontext

Bei der Planung und Durchführung einer wissenschaftlich fundierten Intervention zur Prävention von Vorurteilen und Diskriminierung sollte der soziale Kontext, vor allem in Bezug auf die bestehenden intergruppalen Beziehungen und Einstellungen beachtet werden (Bar-Tal & Teichman, 2005). So wurde bei PARTS auf Grund zweier Pilotstudien in Thüringen (Nitsche, 2008; Küchel, 2008) die Gruppe der Russen als Zielfremdgruppe zur Bearbeitung im Programm gewählt. Hier fanden sich die problematischsten Einstellungen der Studienteilnehmer.

Vorbildfunktion

Durch die Einführung und Beachtung von einfachen Verhaltensregeln zum respektvollen Umgang miteinander, sowie durch Elemente wie den PARTS-Vertrag (s. Anhang A), bei dem solche Regeln explizit festgelegt wurden, zeigten sich die

PARTS-Trainer als Vorbilder für einen respektvollen und offenen Umgang miteinander. Somit setzten sie positive Normen für soziale Einstellungen und Verhaltensweisen.

Insgesamt verfügt das in der vorliegenden Arbeit zu evaluierende Präventionsprogramm PARTS über alle Eigenschaften, die ein wissenschaftlich fundiertes Programm zur Reduktion von Vorurteilen und Diskriminierung mitbringen muss.

V.3. Strukturelle Einordnung von PARTS

Es gibt bisher kaum Evaluationsstudien zur Prävention von Vorurteilen, die eine ähnlich ausführliche Implementationskontrolle wie in der vorliegenden Studie durchgeführt haben, bzw. wenn eine Kontrolle der Implementation durchgeführt wurde, wurde diese nicht in den entsprechenden Studien berichtet. Somit lassen sich die Ergebnisse der vorliegenden Studie zur Implementation des Präventionsprogramms PARTS nicht mit den Ergebnissen bei anderen Studien vergleichen. Daher können lediglich strukturelle Merkmale, wie der inhaltliche Aufbau oder die Intensität und Dauer des Programms, mit anderen Programmen zur Prävention von Vorurteilen und Diskriminierung verglichen werden. Und auch diese Vergleiche sind nur teilweise möglich, da häufig genaue Angaben zu Inhalten, Intensität und Dauer der Interventionen in den Evaluationsstudien fehlen. Im Folgenden werden zunächst die strukturellen Merkmale von PARTS in den Kontext der empirischen Literatur eingeordnet. Dann werden die Ergebnisse der Implementationsanalyse von PARTS zusammengefasst und erörtert.

V.3.1. Inhaltlicher Aufbau des Programms

Wie ausführlich im Methodenteil beschrieben handelt es sich bei PARTS um ein standardisiertes multimodales Trainingsprogramm. Es umfasst drei inhaltlich voneinander abgegrenzte Interventionsbereiche: Stellvertretender-Kontakt-Intervention, wissensbasierte Intervention und die Förderung individueller sozial-kognitiver Kompetenzen. Dabei nehmen die beiden Interventionsbereiche Förderung individueller sozial-kognitiver Kompetenzen und wissensbasierte Intervention deutlich mehr Interventionszeit ein, als der Stellvertretende Kontakt (s.u.).

In der Literatur finden sich im Vergleich dazu meist unimodale Interventionsprogramme (Fisher, 1965; Litcher & Johnson, 1969; Litcher et al., 1973; Yawkey, 1973; Whetstone, 1992; Weiner & Wright, 1973; Johnson et al., 1994; Johnson et al., 1995), oder Studien, die verschiedene Ausführungen eines Interventionsansatzes gegeneinander testen (Cameron & Rutland, 2006; Cameron, Rutland, Brown & Douch, 2006; Howell, 1973; Katz, 1973; Bigler & Liben, 1992). Nur wenige Studien vergleichen die Wirksamkeit komplett unterschiedlicher Interventionsansätze (Cameron, Rutland & Brown, 2007; Katz & Zalk, 1978). Noch

seltener sind echte multimodale Interventionsprogramme. Hier findet sich lediglich das Programm „More Than Meets the Eye“ von Bowers und Swanson (1988), das von Aboud und Fenwick (1999) evaluiert wurde. Dieses Programm beinhaltet hauptsächlich sozial-kognitive Elemente zur differenzierten Wahrnehmung von Personen, aber ebenso wissensbasierte Anteile zur Aufklärung über Vorurteile und Diskriminierung, als auch Selbsterfahrungsberichte über den Umgang mit eigenen oder erlebten Vorurteilen. Das Programm verfügt somit über verschiedene Interventionsansätze, hat aber ähnlich wie PARTS einen sozial-kognitiven Fokus.

Die oben beschriebene Aufteilung der in der Literatur bekannten Interventionsprogramme zur Prävention von gruppenbezogenen sozialen Einstellungsverzerrungen ist vermutlich auch dem Stand der Forschung auf diesem Gebiet geschuldet. Da es bisher noch sehr wenige hochwertige Evaluationsstudien zu diesem Bereich gibt, und so gut wie keine Studien über die zu Grunde liegenden Prozesse von Einstellungsveränderungen, liegt ein grundlagenorientiertes Forschungsvorgehen nahe. Hier wird also zunächst versucht, wirksame Interventionen und Prozesse zu identifizieren, indem unimodale Trainingsprogramme auf ihre Wirksamkeit getestet werden. PARTS hingegen baut auf den Erkenntnissen der erfolgreichen Evaluationsstudien auf und versucht gezielt einen neuen multimodalen Interventionsweg einzuschlagen. Dieses Vorgehen birgt zwar den Nachteil, nicht zweifelsfrei identifizieren zu können, welcher Interventionsbestandteil für eventuelle Einstellungsveränderungen verantwortlich ist. Allerdings ermöglicht es dafür verschiedene als effektiv identifizierte Methoden miteinander kombiniert einzusetzen, und somit ganzheitlich auf verschiedene Facetten von gruppenbezogenen sozialen Einstellungen seine Wirksamkeit zu entfalten (vgl. Dovidio et al., 2004). Weiterhin hat sich gezeigt, dass eine multimodale Vorgehensweise bei Interventionen im Durchschnitt durch eine höhere Wirksamkeit im Vergleich zu unimodalen Interventionen gekennzeichnet ist (Beelmann, 2006).

Somit verfügt PARTS durchaus über eine Art Pinoniercharakter. Mit der Entwicklung und Umsetzung von PARTS stößt die vorliegende Studie in eine Lücke anwendungsbezogener Forschung mit dem Ziel, eine Intervention zu evaluieren, die eine umfassende dauerhafte Veränderung in den gruppenbezogenen sozialen Einstellungen der Kinder ermöglicht.

V.3.2. Intensität und Dauer des Programms

Wie ausführlich im Methodenteil beschrieben (s. Kapitel III. Methoden) erstreckte sich PARTS über 14 Sitzungen mit 15 Interventionseinheiten (die ersten beiden Einheiten waren eine Doppelstunde) über einen Zeitraum von neun Wochen. Dabei hatte jede Interventionseinheit die Dauer einer Schulstunde, also 45 Minuten. Zieht man die zwei Wochen der Herbstferien, die im Programmzeitraum lagen, ohne Interventionseinheiten ab, kommt man auf gut 96 Minuten Interventionszeit pro Woche über 7 Wochen. Im Vergleich dazu liegen die in der Literatur berichteten Interventionen, was Dauer und Intensität angeht, etwas niedriger. So liegt die durchschnittliche Interventionsdauer der im Theorieteil präsentierten empirischen Studien (so denn die Dauer der Intervention in den entsprechenden Studien berichtet wurde) bei 6 Wochen. Dabei erstrecken sich die unterschiedlichen Interventionen über einen Zeitraum von einem Nachmittag (Katz, 1973; Katz & Zalk, 1978) bis hin zu vier Monaten (Litcher & Johnson, 1969). Genauso unterschiedlich ist die Intensität der verwendeten Interventionen. Hier reicht die Bandbreite von einmalig zwei Minuten Diskussion (Aboud & Fenwick, 1999) bis hin zu 18 mal 30 Minuten Training über sechs Wochen (Johnson et al., 1995). Im Durchschnitt liegt die Intensität der Interventionen jedoch bei etwa 50 Minuten Interventionszeit pro Woche. Damit zeigt sich das Gesamtprogramm PARTS, sowohl was die Dauer als auch was die Intensität anbelangt, den durchschnittlichen Interventionsprogrammen aus der Literatur überlegen. Allerdings muss hier festgehalten werden, dass PARTS ein multimodales Programm ist und die Interventionsressourcen auf verschiedene Interventionsbereiche aufgeteilt wurden. Dies ist bei einem Großteil der Trainingsprogramme aus der Literatur nicht der Fall. Vergleicht man nun die Netto-Interventionszeiten der einzelnen Programmbestandteile von PARTS mit den Angaben aus den im Theorieteil berichteten empirischen Studien, zeichnet sich ein anderes, im Folgenden beschriebenes Bild ab.

Stellvertretender-Kontakt-Intervention

Bei PARTS wurden Stellvertretender-Kontakt-Interventionen über 7 Wochen in einem Umfang von 85 Minuten Netto-Interventionszeit mit den Kindern durchgeführt. Das bedeutet im Durchschnitt bekamen die Kinder gut 12 Minuten SK-Intervention pro Woche. In den Studien aus der Arbeitsgruppe von Cameron und Kollegen hingegen fanden die Stellvertretender-Kontakt-Interventionen über sechs Wochen

verteilt für insgesamt 120 Minuten statt. Somit erhielten die Kinder hier 20 Minuten Training pro Woche. Damit zeigt sich hier rein zeitlich eine höhere Interventionsintensität in den Studien der britischen Arbeitsgruppe ab. Zusätzlich wurden die Interventionseinheiten bei den britischen Studien mit Kleingruppen von 2 bis 3 Kindern durchgeführt und nicht wie bei PARTS mit einem kompletten Klassenverband von durchschnittlich 20 Kindern. Die Durchführung in Kleingruppen ermöglicht natürlich eine deutlich intensivere Beteiligung der einzelnen Kinder beim Lesen und Diskutieren der Freundschaftsgeschichten. Insgesamt wurden die Stellvertretender-Kontakt-Interventionen in der Literatur also deutlich intensiver durchgeführt, als die Stellvertretender-Kontakt-Interventionen bei PARTS.

Wissensbasierte Interventionen

In diesen Interventionsbereich wurden bei PARTS 180 Minuten Interventionszeit investiert. Das entspricht bei 7 Wochen Interventionszeit durchschnittlich circa 25 Minuten pro Interventionswoche. Leider finden sich in der Literatur kaum genaue Zeitangaben zur Intensität der durchgeführten wissensbasierten Interventionen. Einzig die Anzahl gelesener Geschichten oder Bücher zum Thema wird genannt (Fisher, 1965; Yawkey, 1973; Howell, 1973), oder die Anzahl der verwendeten Einheiten (Whetstone, 1992) ohne dabei die genaue Dauer der Interventionssitzungen zu beschreiben. Somit kann keine klare Aussage über die Intensität des wissensbasierten Interventionsbestandteils von PARTS im Vergleich mit anderen wissensbasierten Interventionen getroffen werden. Es finden sich aber einige Angaben über die Gesamtdauer der Interventionen aus den empirischen Studien. Diese erstrecken sich im Durchschnitt über einen Zeitraum von 10 Wochen. Damit ist die Interventionsdauer bei PARTS deutlich geringer als bei der durchschnittlichen wissensbasierten Intervention aus der Literatur.

Förderung individueller Kompetenzen

Die Förderung von sozial-kognitiven Fähigkeiten wurde bei PARTS über 200 Minuten²⁰ hinweg mit den Kindern durchgeführt. Das ergibt knappe 30 Minuten Training individueller Kompetenzen pro Interventionswoche. Dabei teilt sich die Förderung der individuellen Fähigkeiten noch weiter auf: Fähigkeit zur multiplen Klassifikation 70 Minuten Gesamtzeit bzw. 10 Minuten pro Woche, Perspektivenübernahmefähigkeit 60 Minuten Gesamtzeit oder gute 8 Minuten pro Woche und soziales Problemlösen mit 70 Minuten Gesamtzeit oder abermals 10 Minuten pro Woche. In den entsprechenden Studien aus der Literatur zur Förderung individueller Kompetenzen wird deutlich, dass durchschnittlich etwa 37 Minuten pro Woche an Interventionszeit verwendet werden; dies allerdings nur über einen Zeitraum von durchschnittlich 4 Wochen, was einer Gesamtdauer von 148 Minuten entspricht. Somit wird rein zeittechnisch bei PARTS mehr Interventionsszeit auf die Förderung individueller Kompetenzen verwendet, als in durchschnittlichen anderen Programmen. Allerdings muss dabei beachtet werden, dass diese 148 Minuten meist nur für das Training einer einzigen Kompetenz, z.B. der multiplen Klassifikationsfähigkeit verwendet werden. Im Gegensatz dazu werden bei PARTS in diesem Bereich lediglich 70 Minuten Interventionszeit eingesetzt. Auch hier zeigt sich also zumindest auf der Ebene der einzelnen zu trainierenden Fähigkeiten, im Vergleich zum Durchschnitt der Interventionen zur Förderung individueller Kompetenzen, eine geringere Intensität dieses Interventionsbestandteils von PARTS.

²⁰ Dabei ist das spielerische Eröffnungsritual „Alle in einer Reihe...“ nicht mit einberechnet, da der Zusammenhang zur multiplen Klassifikationsfähigkeit doch eher gering erscheint.

V.4. Zusammenfassung und Erörterung der Implementationsanalyse

Da es in den bisher veröffentlichten Studien zur Evaluation von Vorurteilspräventionsprogrammen keine Angaben über eine eventuell durchgeführte Implementationskontrolle gibt, können zur Beurteilung des Implementationserfolgs der vorliegenden Arbeit nur die hier verwendeten Skalen des On-Task-Verhaltens, des Off-Task-Verhaltens sowie des Spaß an den Interventionseinheiten und ihre Wertebereiche als Referenzpunkte dienen. Dies zeigt einen eklatanten Mangel an hochwertigen Evaluationsstudien, welche durch die Kontrolle der Implementation die Qualität der zu beurteilenden Intervention sicherstellen.

V.4.1. Implementation allgemein

Die Analysen zur Implementation von PARTS zeichnen insgesamt ein sehr positives Bild. So zeigten die Kinder über das gesamte Programm hinweg ein sehr hohes On-Task-Verhalten. Das bedeutet, sie beteiligten sich aktiv an der Umsetzung des Programms, brachten sich selbst oft und gerne ein und interessierten sich sehr für die dargebotenen Inhalte und Diskussionen. Weiterhin zeigten die Teilnehmer ein sehr niedriges Off-Task-Verhalten über alle Sitzungen hinweg. Dies bedeutet, dass die Kinder den Ablauf der Interventionssitzungen selten und wenig störten, wenig unangemessenes und nicht auf das Programm bezogenes Verhalten zeigten und auch wenig träumten oder unaufmerksam waren. Schließlich zeigten die Teilnehmer auch sehr viel Spaß und Freude an den Interventionseinheiten. Die Spaß-Werte haben bis zum Ende der Intervention auf einem sehr hohen Niveau gelegen, was dafür spricht, dass die Kinder auch in den späteren Interventionseinheiten immer noch mit Freude an den Spielen und Übungen von PARTS teilgenommen haben.

Diese Ergebnisse treffen sowohl bei den Lehrereinschätzungen als auch bei den Trainereinschätzungen zu. Allerdings sind die Korrelationen der Lehrer- und Trainereinschätzungen zu den einzelnen Interventionseinheiten zwar positiv, liegen aber eher im kleinen bis mittleren Größenbereich. Dies spricht für eine unabhängige Beurteilung der Teilnahmequalität der einzelnen Kinder durch Lehrer und Trainer. Weiterhin sind die eher geringen Korrelationen zum Teil sicherlich auch der

Tatsache geschuldet, dass die Lehrer ihre Schüler und deren Unterrichtsverhalten schon lange kennen, während die Trainer ihre Teilnehmer erst einmal kennenlernen mussten, um zuverlässige Einschätzungen ihres Verhaltens treffen zu können.

Insgesamt zeigt sich bei den Durchschnittswerten der Implementationsparameter über alle Sitzungen hinweg jedoch eine deutliche Übereinstimmung zwischen Lehrer- und Trainereinschätzungen. Dies spricht für einen validen positiven Gesamteindruck von der Umsetzung von PARTS.

Die Ergebnisse der Implementationsanalyse der vorliegenden Studie sprechen also für eine erfolgreiche Umsetzung und weiterhin für die Altersangemessenheit des Präventionsprogramms PARTS. So zeigen das hohe On-Task-Verhalten und der hohe Spaß der Kinder an den Interventionseinheiten, dass die Kinder bei der Durchführung des Programms hoch motiviert und konzentriert an deren Umsetzung beteiligt waren. Außerdem können die hohen Werte dieser beiden Implementationsparameter als Hinweis dafür gewertet werden, dass die in PARTS enthaltenen Übungen und Spiele einen hohen Aufforderungscharakter haben, das Mitmachen erleichtern und die Motivation der Kinder dauerhaft durch interessante und spannende Aufgaben gebunden wird.

Das sehr geringe Off-Task-Verhalten der PARTS-Teilnehmer kann als Hinweis für die Alters- und Entwicklungsangemessenheit der Aufgaben des Trainingsprogramms dienen. Bei Aufgaben mit zu hohen oder zu niedrigen Anforderungen werden Kinder schnell frustriert und beginnen, sich mit anderen Dingen zu beschäftigen. Da dies bei der Durchführung von PARTS nicht der Fall war, kann gefolgert werden, dass das Anforderungsniveau der einzelnen Übungen dem Fähigkeitslevel und dem Entwicklungsstand der teilnehmenden Kinder entsprach.

V.4.2. Implementation einzelner Interventionsbereiche

Wenn man die Interventionsbereiche getrennt betrachtet, entsteht ein ähnlich positiver Eindruck wie bei der Implementation des gesamten Programms. Bei allen Interventionsbereichen zeigen sich ein sehr hohes On-Task-Verhalten, also eine hohe Beteiligung der Kinder am Programm, ein sehr niedriges Off-Task-Verhalten, also nur sehr geringes Störverhalten oder Unaufmerksamkeiten, und ein großer Spaß bei allen Programmteilen. Allerdings finden sich auf diesem hohen Niveau auch einige Unterschiede der Implementationsparameter zwischen den einzelnen Programmteilen.

So fallen beim On-Task-Verhalten die Sitzungen zu sozial-kognitiven Kompetenzen leicht gegen die anderen beiden Bereiche ab. In der Lehrereinschätzung wird dieser Unterschied im Vergleich mit dem Wissensinterventionsbereich auch statistisch bedeutsam. Die Einheiten zum interkulturellen Wissen hingegen zeigen die höchsten Werte, und scheinen besonders das Interesse und die Beteiligung der Kinder angeregt zu haben.

Das Off-Task-Verhalten hingegen ist in allen drei Interventionsbereichen etwa gleich und auf einem sehr niedrigen Niveau. Die Kinder haben also unabhängig von den Inhalten der Interventionseinheiten gleich wenig Störverhalten und Aufmerksamkeitsdefizite gezeigt. Hier lässt sich einzig eine, im Vergleich zur Trainereinschätzung, durchgehend höhere Lehrereinschätzung feststellen.

Der Implementationsparameter Spaß an der Interventionseinheit liegt auch in allen drei Bereichen auf einem sehr hohen Level. Nur im Bereich Förderung sozial-kognitiver Fähigkeiten liegen die Werte etwas niedriger. Dieser Unterschied wird in der Lehrereinschätzung gegenüber den beiden anderen Interventionsbereichen auch signifikant. Zudem schätzten die Trainer den Spaß der Kinder bei Wissensinterventionseinheiten signifikant höher ein, als in den anderen beiden Bereichen. Dies spricht dafür, dass die Kinder besonders viel Freude an den Wissensinterventionseinheiten hatten und von den Einheiten zu sozial-kognitiven Kompetenzen nicht ganz so begeistert waren wie von den anderen Programmteilen. Auch wenn die beschriebenen Unterschiede nur sehr kleine Effektstärken zeigen, liefern diese ungeachtet dessen wichtige Hinweise für die Aufnahme der Programminhalte bei den Adressaten. Demnach sind die Einheiten zu interkulturellem Wissen besonders gut aufgenommen worden und die Einheiten, die ausschließlich Übungen zur Förderung sozial-kognitiver Kompetenzen beinhalten, nicht ganz so gut wie der Rest des Programms. Die Einheiten mit stellvertretendem Kontakt spiegeln die mittleren Werte der Implementationsparameter wider. Kein Interventionsbereich zeichnete sich durch erhöhtes Off-Task-Verhalten aus.

Die differentielle Implementationsanalyse der einzelnen Programmbestandteile geben also Hinweise darauf, dass besonders die Interventionen zum interkulturellen Wissen mit sehr guter Beteiligung der Kinder aufgenommen wurden und die Kinder begeisterten. Die Programmeinheiten, die sich ausschließlich mit der Förderung von individuellen sozial-kognitiven Kompetenzen befassten, wurden hingegen etwas weniger positiv angenommen. Das bedeutet,

dass die Kinder die Inhalte der Wissensintervention aktiver aufgenommen und verarbeitet haben und die Inhalte der sozial-kognitiven Förderungen mit weniger Engagement bearbeiteten. Dies zeigt sich auch an den Ergebnissen der Wirksamkeitsanalyse. Hier finden sich deutliche Zuwächse in der Interventionsgruppe beim interkulturellen Wissen, jedoch nur sehr kleine Effekte auf den geförderten sozial-kognitiven Bereichen. Da sich das Off-Task-Verhalten jedoch bei allen Interventionsbereichen stabil auf einem äußerst niedrigen Niveau gezeigt hat, lässt sich daraus schließen, dass die Übungen aller Interventionsmethoden vom Schwierigkeitslevel für das Fähigkeitsniveau der Kinder angemessen sind (s.o.).

V.4.3. Identifikation der Störenfriede

Eine kleine Gruppe von Kindern mit dauerhaft, also über mindestens die Hälfte aller Interventionssitzungen hinweg, erhöhtem Off-Task-Verhalten („Störenfriede“, $n = 26$) konnte identifiziert werden. Allerdings sind auch hier die Einschätzungen von Lehrern und Trainern nicht voll übereinstimmend. Lediglich fünf Fälle sind sowohl von Lehrern als auch von Trainern als Störenfriede identifiziert worden.

Dass der Anteil aller Störenfriede, also der dauerhaft störenden Kindern, recht gering ist und lediglich bei 12.4 % der PARTS-Teilnehmer liegt, spricht ebenfalls für die Altersangemessenheit des Interventionsprogramms und dafür, dass der Großteil der Kinder mit dem Programm gut erreicht werden konnte. Das Trainingsprogramm ist also altersentsprechend angemessen aufgebaut und umgesetzt worden.

V.4.4. Implementation abhängig von Personenmerkmalen

Bei den übrigen möglichen Einflussfaktoren auf die Implementation zeigten sich zu erwartende Effekte der Kovariaten Geschlecht, Alter, Entwicklungskontext und Intelligenz.

Geschlecht

Bei Mädchen konnte das Programm etwas besser als bei Jungen implementiert werden. Dies zeigte sich auf allen drei Implementationsparametern. Es fanden sich signifikant höheres On-Task-Verhalten und signifikant höherer Spaß bei den Mädchen als bei den Jungen. Gleichzeitig zeigten die Mädchen signifikant weniger Off-Task-Verhalten als die Jungen. Diese statistischen Unterschiede sind

auch praktisch nicht zu vernachlässigen, was die kleinen bis mittleren Effektstärken aufzeigen. Mädchen engagierten sich also stärker bei der Umsetzung des Programms, zeigten weniger Störverhalten und mehr Spaß an den Interventionseinheiten als Jungen. Dieses Muster bestätigte sich auch bei der Betrachtung der einzelnen Interventionsbereiche. Insgesamt konnte das Programm bei den Mädchen also noch etwas besser implementiert werden als bei den Jungen. Allerdings sind die Implementationsparameter der Jungen immer noch als sehr gut einzuschätzen. Das Programm lässt sich also trotz der Geschlechtsunterschiede sowohl bei Jungen als auch bei Mädchen sehr gut implementieren.

Alter

Sowohl in den Lehrer- als auch in den Trainereinschätzungen zeigt sich der gleiche Trend: Je älter die Kinder werden, desto schwieriger erweist sich die Implementation des Programms. Dies wird an den negativen Korrelationen zwischen dem Alter der Kinder und dem On-Task-Verhalten sowie dem Spaß deutlich; ebenso wie an der positiven Korrelation zwischen dem Alter und dem Off-Task-Verhalten der Kinder. Das bedeutet, dass jüngere Kinder sich stärker am Programm beteiligten, weniger störten und mehr Spaß bei der Umsetzung von PARTS hatten. Allerdings sind die Zusammenhänge sehr klein und somit praktisch zu vernachlässigen und auch nur bei den Lehrereinschätzungen statistisch bedeutsam. Der gleiche Trend der Zusammenhänge zwischen Alter und den Implementationsparametern zeigt sich auch für die einzelnen Interventionsbereiche. Für Kinder in der dritten Klasse scheint das Programm altersmäßig insgesamt gut geeignet zu sein. Die vorgefundenen Zusammenhänge können aber als Hinweis gewertet werden, dass der Zeitpunkt für die Durchführung des Programms gut gewählt ist, und dass das Programm nicht unbedingt mit einer älteren Gruppe durchgeführt werden sollte.

Entwicklungskontext

Die „Stadtkinder“²¹ zeigen bei Lehrer- und Trainereinschätzungen ein leicht höheres On-Task-Verhalten und zumindest in den Lehrereinschätzungen ein

²¹ Zu der Einteilung von Stadt- und Landkindern muss festgehalten werden, dass die Anzahl der Stadtkinder ($n = 48$), die an PARTS teilgenommen haben deutlich geringer ist, als die Anzahl der Landkinder ($n = 161$). Außerdem gehen die Stadtkinder in drei Thüringer Kleinstädten zur Schule, deren Einwohnerzahlen zwischen 18.000 und 45.000 liegen. Somit ist die Varianz zwischen den Entwicklungskontexten eventuell weniger groß, als innerhalb der einzelnen Kontexte.

signifikant niedrigeres Off-Task-Verhalten als ihre Kameraden aus den ländlichen Schulen. Das bedeutet also, dass die „Stadtkinder“ sich etwas mehr bei der Programmumsetzung engagierten und nach der Lehrereinschätzung etwas weniger störten als die „Landkinder“. Der Unterschied im Off-Task-Verhalten zwischen Stadt- und Landkindern sollte allerdings mit Vorsicht interpretiert werden, da sich dieser Unterschied nur bei den Lehrereinschätzungen findet und die statistischen Voraussetzungen für die verwendete Varianzanalyse bei gleichzeitig ungleich besetzten Zellen hier nicht gegeben waren – der Test also progressiv reagierte. Daher sollte dieser Unterschied nicht überbewertet werden.

Der Spaß der Kinder wird in beiden Entwicklungskontexten in etwa gleich hoch eingeschätzt. Sowohl Stadtkinder als auch Landkinder erfreuten sich gleichermaßen an den Spielen und Übungen von PARTS. Ein ähnliches Muster der Mittelwerte der Implementationsparameter findet sich auch, wenn die drei verschiedenen Interventionsbereiche einzeln betrachtet werden. Insgesamt kann festgehalten werden, dass sich das Programm trotz der kleinen Unterschiede zwischen Stadt- und Landkindern sowohl in der Stadt als auch auf dem Land sehr gut implementieren ließ.

Intelligenz

Kinder, die ein besseres Ergebnis im Raven Standardmatrizentest erreichten, zeigten weniger Off-Task-Verhalten, zeigten also etwas weniger unaufmerksames und störendes Verhalten während des Programms. Dies wird durch die kleinen negativen aber signifikanten Korrelationen zwischen dem Off-Task-Verhalten und dem Ergebnis des Raven Standardmatrizentests deutlich. Die anderen beiden Implementationsparameter On-Task-Verhalten und Spaß zeigen keine Zusammenhänge mit der Intelligenz der Kinder. Das Engagement der Kinder bei der Umsetzung und die Freude an den durchgeführten Übungen sind also nicht abhängig von der Intelligenz der Kinder. Dieses Korrelationsmuster zeigt sich auch bei der getrennten Betrachtung der einzelnen Interventionsbereiche.

Obwohl Kinder mit geringeren Werten im Matrizentest also etwas mehr unaufmerksames und störendes Verhalten während des Programms zeigen, scheint sich dieses dennoch für Kinder mit einer recht großen Bandbreite intellektueller Leistungsfähigkeiten gut zu eignen. Denn zum einen zeigen die anderen Implementationsparameter keine Zusammenhänge zum verwendeten

Intelligenzmaß, und zum anderen zeigen auch die Kinder, die der Risikogruppe 25 des Matrizentests zuzuordnen sind, im Mittel nur ein leicht erhöhtes, immer noch auf einem sehr geringen Niveau liegendes Off-Task-Verhalten ($M = 0.60$ in der Lehrereinschätzung und $M = 0.44$ in der Trainereinschätzung).

Störenfriede

Die Implementation bei den sogenannten „Störenfrieden“ zeigt neben den genuinen Unterschieden zur Gesamtgruppe im Off-Task-Verhalten auch Schwierigkeiten auf den anderen beiden Implementationsparametern On-Task-Verhalten und Spaß. In beiden Bereichen zeigen die Störenfriede geringere Werte, als der Rest der Teilnehmer. Die Störenfriede zeigen also neben dem stark erhöhten Störverhalten auch weniger Beteiligung an den Programminhalten und weniger Spaß an den einzelnen Übungen und Aufgaben des Programms. Allerdings bleiben diese Unterschiede auf einem kleinen bis mittleren Niveau und werden bis auf eine kritisch zu bewertende Ausnahme²² in der Lehrereinschätzung des On-Task-Verhaltens nicht statistisch signifikant. Insgesamt zeigen die Störenfriede also trotz der kleinen Unterschiede beim On-Task-Verhalten und Spaß immer noch eine recht gute Beteiligung am Interventionsprogramm und scheinen auch Spaß an den Sitzungen gehabt zu haben. Das deutlich erhöhte Off-Task-Verhalten spricht aber für gewisse Schwierigkeiten, dem Programm dauerhaft und konzentriert folgen zu können.

Einordnung der Implementationsergebnisse im Zusammenhang mit Personenmerkmalen

Insgesamt kann festgehalten werden, dass die betrachteten Personenmerkmale durchaus einen Einfluss auf die Implementation von PARTS haben. Wenn man, die Ergebnisse der Implementation zusammenfassend, die optimale Adressatin von PARTS beschreibt, ergibt sich folgendes Bild: Das Programm lässt sich besonders gut bei Mädchen implementieren, die in der Stadt wohnen, die nicht zu den Störenfrieden gehören und die etwas jünger und intelligenter sind, als ihre Klassenkameraden.

Diese Implementationseffekte auf den geprüften Personenvariablen spiegeln zumeist bekannte Effekte aus der Schulforschung wider. So ist es ein typisches

²² Die nicht erfüllten statistischen Voraussetzungen und eine ungleiche Gruppengröße bei dieser Varianzanalyse führen zu progressiven Entscheidungen des Tests.

Ergebnis, dass sich Mädchen beim Schulunterricht mehr engagieren, weshalb es auch nicht überrascht, dass Mädchen ein höheres On-Task-Verhalten und mehr Spaß am Programm zeigen als Jungen (vgl. Goodenow, 1993). Auch dass Jungen deutlich mehr störendes Verhalten zeigen als Mädchen, ist wohl eher einer allgemeinen Verhaltenstendenz geschuldet (vgl. Furrer & Skinner, 2003), als einer geschlechtsspezifischen Benachteiligung durch die Programminhalte.

Die Alterseffekte auf die Implementation erweisen sich als sehr gering. Hier zeigen die jüngeren Kinder etwas mehr Beteiligung und Begeisterung an dem Programm und etwas weniger Störverhalten, als ihre älteren Klassenkameraden. Dieses Ergebnis könnte einem allgemeinen Alterseffekt zugeordnet werden, wonach jüngere Kinder sich stärker mit ihrem Umfeld verbunden fühlen und leichter zu führen sind und sich interessierter an allen Aufgaben der Schule zeigen (Goodenow, 1993; Furrer & Skinner, 2003). Es könnte aber auch ein Hinweis dafür sein, dass sich das Programm mit seinem sozial-kognitiven Fokus in dieser Altersgruppe sehr gut durchführen lässt, die teilnehmenden Kinder aber nicht unbedingt viel älter sein sollten, als die Teilnehmer der PARTS-Stichprobe. Bei älteren Kindern könnten andere Interventionsmethoden erfolgversprechender sein (vgl. Kapitel II.2.2. Entwicklungstheorien; Teichmann & Bar-Tal, 2008).

Die Implementationsunterschiede zwischen Stadt- und Landkindern sind ebenfalls sehr gering ausgeprägt. Hauptsächlich findet sich eine leichte Tendenz, dass Stadtkinder sich mehr am Programm beteiligen als Landkinder. Dies könnte daran liegen, dass die Stadtkinder mehr Kontakt zu Mitgliedern von ethnischen Fremdgruppen haben und diesem Thema allgemein etwas aufgeschlossener gegenüberstehen als Landkinder.

Die Effekte der Intelligenz der Kinder auf die Implementation beschränken sich auf einen kleinen negativen Zusammenhang zwischen der Intelligenz und dem Störverhalten der Kinder. Je intelligenter, desto weniger Störverhalten zeigen die Teilnehmer. Auch dies ist ein zu erwartender Effekt, da intelligentere Kinder zum einen den Programminhalten leichter folgen können und weniger Schwierigkeiten mit der Konzentration haben, zum anderen aber auch die sozialen Normen zum ordnungsgemäßen Verhalten im Klassenraum besser gelernt haben könnten.

Die Implementationsparameter bei den Störenfriede(n) liegen insgesamt noch auf einem sehr guten Niveau. Trotz der deutlichen und genuinen Unterschiede im Off-Task-Verhalten, zeigen sich nur kleine bis mittlere Unterschiede in den Bereichen

On-Task-Verhalten und Spaß. Hier zeigen die Störenfriede etwas weniger gute Programmbeteiligung und etwas weniger Spaß an den Einheiten, als ihre übrigen Klassenkameraden. Allerdings liegen die Werte der Störenfriede in allen drei Implementationsbereichen immer noch auf einem recht guten Level. Dies spricht dafür, dass PARTS auch mit einzelnen schwierigeren Kindern immer noch sehr gut umgesetzt werden kann, und dass auch diese Kinder relativ gut am Programm teilnehmen und davon profitieren können.

V.4.5. Fazit zur Implementationsanalyse

All diese Effekte bleiben aber durchweg auf einem kleinen bis maximal mittleren Niveau. Dies spricht sowohl für eine gute Auswahl von angemessenen Aufgaben und Übungen als auch für eine erfolgreiche Umsetzung des Trainingsprogramms in den Klassen. Daher lässt sich zusammenfassend festhalten, dass das Präventionsprogramm PARTS insgesamt sehr gut implementiert werden konnte. Dafür sprechen sowohl die oben zusammengefassten Ergebnisse der differenzierten Implementationsanalyse mit den drei verwendeten Implementationsparametern On-Task-Verhalten, Off-Task-Verhalten und Spaß, als auch die persönlichen Erfahrungen der Trainer, die von den teilnehmenden Kindern zu den einzelnen Sitzungen stets begeistert begrüßt wurden und viel positives Feedback sowohl von den Kindern als auch von den beteiligten Lehrern erhielten.

Insgesamt ließ sich das Präventionsprogramm PARTS sehr gut bei Grundschulkindern der dritten Klasse implementieren.

V.5. Zusammenfassung und Erörterung der Wirksamkeitsanalyse

Im Folgenden werden die Ergebnisse der Wirksamkeitsanalyse zusammengefasst und diskutiert. Da sich bei den Risikogruppen 25 grundsätzlich ähnliche Ergebnismuster wie bei den globalen Auswertungen zeigen, finden sich die Auswertungen und Zusammenfassungen der Risikogruppen gesammelt im Anhang (s. Anhang D).

V.5.1. Zusammenfassung und Einordnung der Ergebnisse zu den proximalen Effektmaßen

Die proximalen Effektmaße bilden die Bereiche ab, welche durch PARTS direkt angesprochen werden. Zum einen das interkulturelle Wissen der Kinder, zum anderen drei sozial-kognitive Kompetenzen: Die multiple Klassifikationsfähigkeit, die Perspektivenübernahmefähigkeit und die soziale Informationsverarbeitung. Schließlich gehört noch der Aufbau eines positiven stellvertretenden Kontakts mit Mitgliedern der russischen Zielfremdgruppe zu den proximalen Zielen von PARTS. Dies wurde jedoch nicht proximal operationalisiert, sondern wird über die Wirkung auf den entsprechenden distalen Einstellungsmaßen sowie über die Implementationsparameter zu den Sitzungen mit Stellvertretender-Kontakt-Interventionen erschlossen. Die Ergebnisse zu diesen proximalen Effektmaßen werden im Folgenden zusammenfassend dargestellt.

Auf einem Großteil der proximalen Effektmaße findet sich insgesamt ein Zuwachs der gemessenen Fähigkeiten.

Interkulturelles Wissen

So zeigt sich beispielsweise im Bereich des interkulturellen Wissens bei allen Kindern ein deutlicher Wissenszuwachs über die Zeit. Dieser Zuwachs ist bei den Teilnehmern der Interventionsgruppe im Vergleich zur Kontrollgruppe deutlich größer. Dies zeigt sich sowohl an den großen positiven Netto-Effektstärken im Prä-Post-Vergleich und im Prä-Follow-up-Vergleich, als auch an der signifikanten Interaktion der Faktoren Messzeitpunkte und Gruppe. Dabei ist der langfristige Effekt

sogar noch größer als der kurzfristige²³. Post-hoc erweisen sich signifikante Entwicklungsunterschiede der beiden Untersuchungsgruppen in allen Messzeitpunktvergleichen.

Das bedeutet, dass die Kinder der Interventionsgruppe sowohl kurzfristig und besonders auch langfristig deutlich mehr interkulturelles Wissen hinzugewonnen haben, als ihre Kameraden aus der Kontrollgruppe.

Einordnung der Ergebnisse zum interkulturellen Wissen

In vielen älteren Studien zu wissensbasierten Interventionen wurden leider keine proximalen Maße zur Prüfung der Intervention verwendet. In den Ergebnissen aktuellerer Studien finden sich jedoch proximale Effektmaße. Hier korrespondieren die Ergebnisse teilweise mit den vorgefundenen Befunden der PARTS Studie. So berichten Perkins und Mebert (2005) von einem erhöhten ethnienpezifischen Wissen (Gemeinsamkeiten und Unterschiede zwischen Ethnien), nachdem eine Gruppe von Vorschülern ein kindbezogenes multikulturelles Curriculum durchlaufen hat. Lauren Whetstone (1992) hingegen konnte bei der Evaluation der „A World of Difference-Materialien“ lediglich einen Wissenszuwachs (Wissen über relevante Begriffe wie Vorurteile und Diskriminierung) bei der Gruppe der Elftklässler verzeichnen, während die Drittklässler und die Gruppe der Siebtklässler hier nicht profitieren konnten. Allerdings zeigten sich unabhängig von der Wirkung auf proximale Effektmaße weder bei Perkins und Mebert (2005) noch bei Whetstone (1992) Effekte der Interventionen auf die gruppenbezogenen sozialen Einstellungen der Kinder. Ob dies auf eine zu geringe Effektstärke beim interkulturellen Wissen zurückgeführt werden kann, muss leider spekuliert werden, da in beiden Studien keine Effektstärken angegeben wurden und auch keine Wirkprozessanalysen durchgeführt wurden.

PARTS zeigt sehr große Effekte beim interkulturellen Wissen. Eine Ursache für die großen Effekte ist sicherlich die Nähe zwischen Intervention und Effektmaß. Der Wissenstest wurde eigens für die vorliegende Studie entwickelt und zumindest teilweise auf die Inhalte der wissensbasierten Intervention von PARTS abgestimmt. Somit war zu erwarten, dass bei einer erfolgreichen Implementation der Wissensintervention auf dem proximalen Effektmaß positive Effekte entstanden.

²³ „Kurzfristig“ meint im Prä-Post-Vergleich, während „langfristig“ sich auf den Prä-Follow-up-Vergleich bezieht.

Weiterhin ist das positive Ergebnis auf dieser Skala ein Hinweis für das Erreichen der zum Wissenszuwachs notwendigen Interventionsdauer und –intensität. Die Vergrößerung des Effekts beim interkulturellen Wissen über die Zeit deutet allerdings darauf hin, dass durch PARTS auch eine Entwicklung angestoßen wurde, der die Teilnehmer nach Beendigung des Programms weiter folgten. Hier scheint auf jeden Fall das Interesse der Kinder für unterschiedliche Kulturen und deren Lebensweisen geweckt worden zu sein.

Multiple Klassifikationsfähigkeit

Auch im Bereich der multiplen Klassifikationsfähigkeit zeigt sich eine positive Entwicklung beider Untersuchungsgruppen über die Zeit. Dabei findet sich nur ein sehr kleiner nicht-signifikanter Vorsprung der Interventionsgruppe gegenüber der Kontrollgruppe. Dies zeigt sich auch an den zwar positiven aber sehr kleinen Netto-Effektstärken zu beiden Vergleichszeiträumen. In der entsprechenden Risikogruppe 25 zeigt sich ein ähnliches Ergebnismuster zwar mit größeren Effektstärken, die allerdings auch nicht statistisch signifikant werden (s. Anhang D).

Insgesamt zeigten Kinder die an PARTS teilnahmen kurzfristig und langfristig eine leicht verbesserte Klassifikationsfähigkeit mit sozialen Stimuli, als die Kinder der Kontrollgruppe.

Einordnung der Ergebnisse zur multiplen Klassifikationsfähigkeit

Im Vergleich zu diesen Ergebnissen, finden sich in der Literatur deutlich größere Effekte bei der Förderung von multipler Klassifikationsfähigkeit. Sowohl in der Studie von Bigler und Liben (1992) als auch in der Studie von Cameron und Kollegen (2007) zeigen sich mittlere bis große statistisch bedeutsame (Netto-) Effektstärken²⁴ nach einem Klassifikationstraining. So zeigt sich bei Bigler und Liben (1992), dass abhängig vom durchgeführten Training (Klassifikationstraining mit sozialen Stimuli, Klassifikationstraining mit nicht-sozialen Stimuli oder Regeltraining) Netto-Effektstärken zwischen $d = 0.90$ und $d = 2.39$ im Prä-Post-Vergleich vorliegen. Wobei das Klassifikationstraining wirkungsvoller war als das Regeltraining. Dies gilt insbesondere beim Training mit sozialen Stimuli. Auch in der Vergleichsstudie von Cameron und Kollegen (2007) zeigte sich ein deutlicher kurzfristiger Effekt eines

²⁴ Auch wenn die Autoren keine Effektstärken zu ihren Ergebnisse angeben, lassen sich diese anhand der Mittelwerte und Standardabweichungen berechnen.

ähnlichen Trainings multipler Klassifikation im Vergleich zur Kontrollgruppe. Die Effektstärke innerhalb des Posttests lag hier bei $d = 0.75$ auf einem ähnlichen Klassifikationsmaß (Bigler & Liben, 1995).

Allerdings müssen diese Effekte von Bigler und Liben sowie Cameron und Kollegen etwas relativiert werden. Der Hauptkritikpunkt dabei ist die große Übereinstimmung zwischen der Trainingsmethode und dem Messinstrument. Die Kinder wurden während des Trainings mit dem gleichen Aufbau an Material trainiert, wie es auch in der Messung verwendet wurde. Somit sind hier große Effektstärken mehr als zu erwarten. Weiterhin wurden die Messungen direkt vor und nach bzw. bei Cameron und Kollegen ausschließlich direkt nach dem Training durchgeführt. Im Vergleich dazu wurde bei PARTS der Posttest erst zwei Monate nach Beendigung des Trainingsprogramms durchgeführt. Über mögliche längerfristige Effekte treffen weder Bigler und Liben noch Cameron und Kollegen Aussagen. Und schließlich war die Intensität und Dauer der Intervention für das reine Klassifikationstraining mit insgesamt gut 120 Minuten Training in einer Woche, bzw. über sechs Wochen verteilt, deutlich intensiver als das Klassifikationstraining bei PARTS. Dies nahm insgesamt nur 70 Minuten in Anspruch. Weiterhin waren die Klassifikationsübungen bei PARTS in keiner Weise auf das Klassifikationsmaß abgestimmt, sondern zielten auf das Trainieren dieser Fähigkeit für den realen Alltag mit echten Personen ab. Und schließlich zeigten die Teilnehmer der PARTS-Studie bereits zum ersten Messzeitpunkt im Durchschnitt deutlich mehr multiple Klassifikationsfähigkeit ($M = 7.69$, $SD = 2.46$), als die Teilnehmer der Studie von Bigler und Liben (1992) ($M = 2.93$, $SD = 3.05$)²⁵. Somit ist der Entwicklungsspielraum der Bigler und Liben-Stichprobe deutlich größer als in der PARTS-Stichprobe. Bei Cameron und Kollegen beruht das Ergebnis ausschließlich auf dem Posttest, da kein Prätest durchgeführt wurde.

Somit ist es nicht verwunderlich, dass in obigen Studien deutlich größere proximale Effekte durch ein Klassifikationstraining erzielt wurden, als dies bei PARTS der Fall war. Dass bei PARTS trotz eines gezielten Trainings allerdings nur kleine trendmäßige Effekte vorliegen, kann verschiedene Gründe haben. Hier scheint vor allem die große Diskrepanz zwischen den Interventionsinhalten und dem Messverfahren, welches an das Verfahren von Bigler und Liben angelehnt ist, Ursache für die kleinen Effekte zu sein. Somit ist die Einordnung des

²⁵ Bei einem Wertebereich von 0-12 auf beiden Messinstrumenten.

Klassifikationstest als proximales Effektmaß nur oberflächlich betrachtet gerechtfertigt. Der Transfer der Interventionsinhalte auf die Anwendung beim Klassifikationstest hat hier nicht gut funktioniert. Weiterhin ist die Intensität des Klassifikationstrainings bei PARTS mit durchschnittlich 10 Minuten pro Woche Netto-Interventionszeit relativ gering gewesen. Unter diesen Umständen ist es insgesamt betrachtet, durchaus als Erfolg zu werten, dass mit einer validen relativ distalen Messmethode kurzfristige positive Effekte der Intervention erfolgten und diese auch längerfristig erhalten blieben.

Perspektivenübernahmefähigkeit

Im Bereich der Perspektivenübernahme liegen die Werte des Chandler-Cartoon-Tests zum ersten Messzeitpunkt in beiden Untersuchungsgruppen relativ hoch. Es findet sich daraufhin eine leichte positive Entwicklung der Werte in beiden Untersuchungsgruppen, wobei die Werte der Interventionsgruppe über die Zeit etwas mehr steigen als die der Kontrollgruppe. Dies zeigt sich an den kleinen positiven Netto-Effektstärken zu beiden Vergleichszeiträumen. Allerdings werden diese kleinen Effekte statistisch nicht bedeutsam.

Insgesamt bedeutet das, dass die Kinder beider Untersuchungsgruppen sich in ihrer Perspektivenübernahmefähigkeit über die Zeit leicht verbessern. Dabei nimmt die Interventionsgruppe nur tendenziell etwas mehr in ihrer Fähigkeit zu als die Kontrollgruppe.

Einordnung der Ergebnisse zur Perspektivenübernahmefähigkeit

In der Literatur finden sich nur sehr wenige Studien zur Förderung von Perspektivenübernahme und Prävention von Vorurteilen. So konnten Weiner und Wright (1973) zwar zeigen, dass das Interventionsprogramm Blue Eyed/Brown Eyed wirksam bei der Veränderung von gruppenbezogenen sozialen Einstellungen gegenüber Schwarzen ist, verwendeten aber leider keine proximalen Effektmaße um zu prüfen, ob diese Einstellungsveränderungen durch eine erhöhte Empathie oder Perspektivenübernahme gegenüber Diskriminierungsopfern erzeugt wurde.

In einer Interventionsstudie von Chandler (1973) zur Reduktion von delinquentem Verhalten finden sich jedoch proximale Effektmaße. Chandler trainierte chronisch delinquente Jungen im Alter zwischen 11 und 13 Jahren mittels einer Intervention, basierend auf Schauspieltraining und Videodrehs. So sollten die Jungen

üben, sich in andere Rollen hineinzusetzen und sich selbst aus anderen Perspektiven wahrzunehmen. Die über zehn Wochen durchgeführte Intervention beinhaltete eine dreistündige Sitzung pro Woche. Die Ergebnisse zeigten eine signifikante Verbesserung der Trainingsgruppe in ihrer Perspektivenübernahmefähigkeit zwischen Prä- und Posttest im Vergleich zur Kontrollgruppe. Leider wurden hier weder Effektstärken berechnet, noch alle nötigen statistischen Kennwerte zu deren nachträglicher Berechnung angegeben.

Die tendentiellen kleinen positiven Effekte auf der Perspektivenübernahmefähigkeit nach der Teilnahme an PARTS, im Vergleich zu dem signifikanten Zuwachs an Perspektivenübernahmefähigkeit bei Chandler, lassen sich durch drei Faktoren erklären. Zum einen war die Intensität und Dauer des Trainingsprogramms in der Chandler-Studie mit 30 Stunden Interventionszeit über 10 Wochen, bei Kleingruppen von fünf Jungen pro Trainingsgruppe, sehr hoch. Im Vergleich dazu wurden bei PARTS nur 60 Minuten über 7 Wochen an Perspektivenübernahmetrainings im Klassenverband durchgeführt. Dies entspricht gerade mal 8 Minuten pro Woche. Hier scheint die Intensität des Trainings an der Perspektivenübernahmefähigkeit nicht hoch genug gewesen zu sein.

Zum anderen zeigten die Teilnehmer an der PARTS-Studie bereits zum ersten Messzeitpunkt ein relativ hohes Niveau in ihrer Perspektivenübernahmefähigkeit. Somit war der Spielraum für eine positive Entwicklung hier wesentlich geringer, als in der chronisch delinquenten Trainingsgruppe von Chandler, die zur Prätestmessung schlechtere Werte als ihre „normale“ Kontrollgruppe aufwies. Korrespondierend zu dieser Erklärung findet sich in der Risikogruppe 25 zumindest verzögert eine mittelgroße positive Netto-Effektstärke nach dem PARTS-Training (s. Ahnang D).

Ein weiterer Grund, warum das Training der Perspektivenübernahmefähigkeit bei PARTS weniger erfolgreich erscheint, liegt vermutlich auch in der geringen internen Konsistenz des Messinstruments. Auch wenn Chandler (1973) das gleiche Instrument verwendete, nutzte er eine wesentlich ausführlichere Variante mit 10 Cartoons und erzielte damit eine split-half-Reliabilität von $\alpha = .92$. Bei der vorliegenden PARTS-Studie wurde aus Gründen des Umfangs und der Belastung der Kinder durch eine Vielzahl von Tests und Interviews das Instrument auf 2 Geschichten gekürzt. Daraus resultierte jedoch eine sehr geringe interne Konsistenz zwischen $\alpha = .28$ und $\alpha = .45$. Dies ist ein Indikator für sehr viel Störvarianz auf dem

Messinstrument, was die Identifikation von möglichen Evaluationseffekten deutlich erschwert und überlagert.

Und schließlich hängen hier Interventionsinhalte und Messmethode bei PARTS nicht zusammen, so dass an dieser Stelle nicht wirklich von einem proximalen Effektmaß ausgegangen werden kann²⁶.

Somit sind die kleinen positiven Netto-Effektstärken, die sich sowohl kurzfristig als auch langfristig im Vergleich von Interventionsgruppe und Kontrollgruppe zeigen, durchaus als Erfolg für PARTS zu werten.

Soziale Informationsverarbeitung

Die Werte des Tests zum sozialen Problemlösen liegen in beiden Untersuchungsgruppen bereits zum ersten Messzeitpunkt auf einem relativ hohen Niveau. Über die Zeit hinweg bleiben diese Werte in beiden Gruppen sehr stabil. Es zeigen sich also weder in der Interventions- noch in der Kontrollgruppe kurz- oder langfristige Veränderungen. Dies wird auch an den nahe Null liegenden Nettoeffektstärken und der nicht bedeutsamen Interaktion der Faktoren Messzeitpunkte und Gruppe deutlich.

Das bedeutet, dass die Fähigkeit zur sozialen Informationsverarbeitung der Kinder im Zeitraum zwischen Prätest und Follow-up in beiden Untersuchungsgruppen stabil bleibt. Nur die Kinder mit anfänglichen Schwierigkeiten der Risikogruppe 25 konnten hier tendentiell kurzfristig von der Intervention profitieren und ihre soziale Informationsverarbeitung schneller entwickeln als die Risikokinder der Kontrollgruppe (s. Anhang D).

Einordnung der Ergebnisse zur sozialen Informationsverarbeitung

In den beiden Studien von Johnson und Kollegen (1994, 1995) zur Evaluation eines Peer-Verhandlungs- und Mediationstrainings hingegen ließ sich das soziale Problemlösen von Dritt- bis Sechstklässlern sehr gut trainieren. In der ersten Studie konnten die Trainer eine signifikante Veränderung der Lösungsstrategien auf einem Konfliktszenariomaß sowohl im Prä-Post-Vergleich als auch im Prä-Follow-up-Vergleich feststellen. Hier wurde von den Kindern vor allem mehr eigenverantwortliches Handeln beschrieben und dabei auf die erlernte Verhandlungs/Mediationsstrategie zurückgegriffen. Ein ähnlich positives Ergebnis

²⁶ Dies gilt auch für die Chandler-Studie (1973).

erbrachte die Folgestudie von Johnson und Kollegen (1995). Hier ließen sich innerhalb des Posttest große positive Effektstärken zwischen $d = 0.76$ und $d = 1.02$ im Vergleich zwischen Interventions- und Kontrollgruppe auf einem Konfliktszenariomaß finden.

Zur Klärung dieser unterschiedlichen Befunde der vorliegenden Arbeit und der Studien von Johnson und Kollegen zur sozialen Informationsverarbeitung lassen sich drei Faktoren heranziehen. Erstens ist die Intensität des Trainings im sozialen Problemlösen bei PARTS mit 70 Minuten über 7 Wochen verteilt, nur sehr gering gewesen. Im Vergleich dazu verwendeten Johnson und Kollegen zwischen 9 und 15 Stunden zur Vermittlung und Übung ihres Verhandlungs- und Mediationstrainings. Somit sind in diesem Bereich bei PARTS auf Grund der geringen Trainingsintensität sowieso nur kleine Interventionseffekte zu erwarten. Zweitens lagen die Prätestwerte in der vorliegenden Studie bereits auf einem relativ hohen Niveau, so dass es weniger Entwicklungsspielraum nach oben gab. Um den vorhandenen Spielraum zu nutzen, hätte es eines intensiveren Trainings bedurft. Und drittens gab es auch in diesem Bereich Probleme mit dem Messinstrument. Auch hier ist das Instrument nicht eng genug auf die Interventionsinhalte von PARTS abgestimmt, um wirklich als proximales Maß eingeordnet zu werden. Und auch der Test zum sozialen Problemlösen weist nur sehr geringe Reliabilitäten zwischen $\alpha = .41$ und $\alpha = .43$ auf. Dies erschwert das Auffinden möglicher kleiner Interventionseffekte zusätzlich.

Insgesamt zeigt diese Skala keine Interventionseffekte des Trainingsprogramms. Die hier gemessenen Anteile sozialer Informationsverarbeitung bleiben über den Zeitraum zwischen Prätest und Follow-up in beiden Untersuchungsgruppen stabil.

Stellvertretender-Kontakt

In der vorliegenden Studie wurde darauf verzichtet, ein proximales Erfolgsmaß für die Herstellung positiven stellvertretenden Kontakts einzusetzen. Dies geschah zum einen, weil aus bisherigen Studien keine Probleme bei der Kontaktherstellung berichtet wurden (vgl. Cameron et al., 2006a, 2006b, 2007; Wright et al., 1997), und zum anderen, weil bereits eine hohe Belastung der Teilnehmer durch eine Vielzahl eingesetzter Instrumente bestand.

Die Erreichung des proximalen Interventionsziels, positiven stellvertretenden Kontakt zwischen den PARTS-Teilnehmern und drei russischen Kindern aus den drei

Freundschaftsgeschichten herzustellen, wird über die Ergebnisse aus der Implementationsanalyse sowie aus den Ergebnissen zu gruppenbezogenen sozialen Einstellungen erschlossen.

Die Implementationsparameter bei Interventionseinheiten mit Stellvertretender-Kontakt-Intervention weisen auf eine erfolgreiche Implementation der Freundschaftsgeschichten hin. Es zeigen sich hier ein sehr hohes On-Task-Verhalten, sehr geringes Off-Task-Verhalten sowie großer Spaß der Kinder während dieser Interventionseinheiten (s.o.). Leider bieten die bisherigen Interventionsstudien zu Stellvertretender-Kontakt hier keine Vergleichsmöglichkeiten, da dort keine Kontrolle der Implementationsparameter vorgenommen wurde. Die gute Implementierbarkeit dieser Lektionen ist jedoch ein Hinweis für die erfolgreiche Herstellung eines stellvertretenden Kontakts durch die verwendeten Freundschaftsgeschichten.

Des Weiteren spricht eine positive Entwicklung bei den gruppenbezogenen sozialen Einstellungen gegenüber der russischen Fremdgruppe in der Interventionsgruppe (s.u.) ebenfalls für ein erfolgreiches Herstellen eines positiven stellvertretenden Kontakts in der Interventionsgruppe. Hier zeigen sich, vor allem verzögert, auf verschiedenen Einstellungsmaßen (mit kognitivem, affektivem und konativem Schwerpunkt) positive Effekte der Intervention. Dies weist deutlich auf eine gelungene Stellvertretender-Kontakt-Intervention hin, da ausschließlich bei dieser Trainingsmethode gezielt auf die russische Fremdgruppe eingegangen wurde und sich bei den Einstellungsmaßen gegenüber anderen Fremdgruppen kaum positive Effekte finden lassen (s.u.).

Fazit zur Wirksamkeitsanalyse – Proximaler Effektmaße

Insgesamt findet also, abgesehen von der sozialen Informationsverarbeitung, ein Zuwachs auf den gemessenen proximalen Fähigkeiten statt. Die Interventionsgruppe kann dabei, gegenüber der Kontrollgruppe, vom Interventionsprogramm profitieren. Dies zeigt sich insbesondere beim interkulturellen Wissen durch große signifikante Interventionseffekte. Aber auch im Bereich der multiplen Klassifikationsfähigkeit und der Perspektivenübernahmefähigkeit zeigen sich, zumindest tendentiell, kleine positive Effekte des Trainingsprogramms. Diese Interventionseffekte sind sowohl kurzfristig als auch langfristig messbar. Nur im Bereich der sozialen Informationsverarbeitung zeigen sich weder eine allgemeine

Veränderung, noch ein bemerkenswerter Effekt des Trainingsprogramms. Die Herstellung eines positiven Stellvertretenden-Kontakts zu russischen Kindern scheint nach den Ergebnissen der Implementationsanalyse sowie der gruppenbezogenen sozialen Einstellungen gegenüber der russischen Fremdgruppe erfolgreich gewesen zu sein. Ähnliche Ergebnismuster finden sich auch in den entsprechenden Risikogruppen 25 der einzelnen proximalen Skalen wieder (s. Anhang D). Die verwendeten Kovariaten zeigen keinen Einfluss auf die oben geschilderten Ergebnisse zu den proximalen Effektmaßen.

Die Ursachen für die nur kleinen oder nicht vorhandenen Interventionseffekte auf den proximalen Effektmaßen werden dabei in der mangelnden Trainingsintensität, den hohen Prätestwerten und den methodischen Problemen der Messinstrumente gesehen.

V.5.2. Zusammenfassung und Einordnung der Ergebnisse zu den distalen Effektmaßen

Zu den distalen Effektmaßen zählen die Maße zu den gruppenbezogenen sozialen Einstellungen mit kognitivem, affektivem und konativem Schwerpunkt. Weiterhin gehören die interpersonale Toleranz und das allgemeine Sozialverhalten zu den distalen Effektmaßen. Die Ergebnisse zu den entsprechenden Analysen werden im Folgenden zusammengefasst.

Gruppenbezogene soziale Einstellungen mit kognitivem Schwerpunkt

Die gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt²⁷ zeichnen sich fast durchgängig durch einen Abfall der Einstellungsverzerrungen über die Zeit in der gesamten Stichprobe aus. Dabei liegen die Werte der Einstellungsverzerrungen auch zum ersten Messzeitpunkt, verglichen mit dem Wertebereich der Skalen, auf einem sehr niedrigen Niveau. Auffällig ist hierbei, dass die Einstellungsverzerrungen gegenüber der chinesischen Fremdgruppe insgesamt noch etwas geringer sind, als die Einstellungsverzerrungen gegenüber der russischen Fremdgruppe. Bei den einzelnen Skalen zu den gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt ließen sich ähnliche Trends und Entwicklungsverläufe feststellen.

Kognitive Einstellungen gegenüber der russischen Zielfremdgruppe

Bei der MRA-Eigengruppenbevorzugung gegenüber einem russischen Kind finden sich signifikant unterschiedliche Entwicklungsverläufe. In der Kontrollgruppe gibt es zunächst einen leichten Abfall der Werte und zum dritten Messzeitpunkt einen Wiederanstieg. In der Interventionsgruppe ist der Verlauf genau umgekehrt. Hier gibt es zum zweiten Messzeitpunkt einen leichten Anstieg, aber zum dritten Messzeitpunkt einen Abfall der Werte. Diese gegenläufigen Entwicklungsverläufe zeigen sich auch in den kleinen, zunächst positiven (prä-post) und dann negativen (prä-follow-up) Netto-Effektstärken. Dabei weisen die post-hoc-Kontraste daraufhin, dass nur zwischen dem zweiten und dritten Messzeitpunkt die signifikante Interaktion der Faktoren Zeit und Gruppe vorliegt.

²⁷ Dazu gehören die Eigengruppenbevorzugungs- und die Fremdgruppenabwertungsskala des Multi-Response Racial Attitude Measure (MRA) sowie die Verkuyten-Skala.

Das heißt, der bedeutsame Gruppenunterschied qualifiziert sich dadurch, dass die Eigengruppenbevorzugung gegenüber einem russischen Kind von t2 zu t3 in der Interventionsgruppe abnimmt, während in der Kontrollgruppe in diesem Zeitraum ein Wiederanstieg der Eigengruppenbevorzugung zu verzeichnen ist. In der entsprechenden Risikogruppe findet sich tendentiell eine ähnlich Entwicklung, die allerdings nicht statistisch signifikant wird (s. Anhang D).

Ein ähnliches Ergebnismuster zeigt sich bei der MRA-Fremdgruppenabwertung gegenüber einem russischen Kind. Auch hier finden sich statistisch bedeutsame unterschiedliche Entwicklungsverläufe beider Untersuchungsgruppen. In der Interventionsgruppe zeigt sich von t1 zu t2 ein leichter Anstieg, von t2 zu t3 aber ein deutlicher Abfall in der Fremdgruppenabwertung. Dagegen bleibt das Niveau der Fremdgruppenabwertung in der Kontrollgruppe eher stabil. Dies führt zu der kleinen negativen Prä-Follow-up-Netto-Effektstärke sowie zu einem statistisch signifikanten post-hoc-Kontrast zwischen dem zweiten und dem dritten Messzeitpunkt.

Das bedeutet, dass die kognitive Fremdgruppenabwertung gegenüber einem russischen Kind der Interventionsgruppe besonders zwischen dem zweiten und dritten Messzeitpunkt signifikant abnimmt, während die Fremdgruppenabwertung in der Kontrollgruppe stabil bleibt oder, wie in der entsprechenden Risikogruppe 25, wieder ansteigt (s. Anhang D).

Gestützt werden diese Ergebnisse durch den Befund auf der Verkuyten-Skala gegenüber der Gruppe russischer Kinder zum dritten Messzeitpunkt. Hier zeigt die Interventionsgruppe signifikant weniger kognitive Einstellungsverzerrungen/Eigengruppenbevorzugung als die Kontrollgruppe. Dies wird auch durch die kleine negative Effektstärke innerhalb des dritten Messzeitpunktes deutlich.

Kognitive Einstellungen gegenüber der chinesischen Fremdgruppe

Bei der MRA-Eigengruppenbevorzugung gegenüber einem chinesischen Kind liegen die Werte zum Prätest bereits auf einem sehr niedrigen Level, wobei die Werte in der Interventionsgruppe noch niedriger liegen als in der Kontrollgruppe. Es findet sich ein Abfall der Werte in beiden Untersuchungsgruppen zwischen den ersten beiden Messzeitpunkten, wobei die Werte der Kontrollgruppe stärker absinken als die der Interventionsgruppe. Dies und der höhere Prätestwert in der

Kontrollgruppe führen zu den sehr kleinen positiven Netto-Effektstärken zu beiden Vergleichszeiträumen. Allerdings fallen die Werte in der Interventionsgruppe auch zum dritten Messzeitpunkt noch weiter, sogar in den negativen Wertebereich, was einer minimalen Fremdgruppenbevorzugung gleichkommt. Währenddessen lässt sich in der Kontrollgruppe ein leichter Wiederanstieg der Werte verzeichnen. Dabei sind die unterschiedlichen Entwicklungen der beiden Gruppen statistisch nicht signifikant.

Das bedeutet, dass die Eigengruppenbevorzugung gegenüber einem chinesischen Kind zunächst in beiden Untersuchungsgruppen absinkt. Allerdings setzt sich dieser Abfall tendentiell nur in der Interventionsgruppe auch zwischen dem zweiten und dritten Messzeitpunkt fort, während es in der Kontrollgruppe einen leichten Wiederanstieg bei der Eigengruppenbevorzugung in diesem Zeitraum gibt. Dabei nähern sich jedoch die Niveaus der beiden Untersuchungsgruppen über die Zeit einander an.

Bei der MRA-Fremdgruppenabwertung gegenüber einem chinesischen Kind sehen die Entwicklungsverläufe sehr ähnlich aus. Auch hier liegen die Werte bereits zum ersten Messzeitpunkt auf einem sehr niedrigen Niveau und fallen über die Zeit in beiden Untersuchungsgruppen weiter ab. Diese Entwicklung verläuft hier allerdings in etwa parallel in beiden Untersuchungsgruppen. Daher finden sich auch nur minimale Netto-Effektstärken in beiden Vergleichszeiträumen. Somit zeigen sich weder deskriptiv noch inferenzstatistisch Effekte der Intervention. Erwähnenswert ist noch die Tatsache, dass es auch auf dieser Skala einen tendentiellen Niveauunterschied zwischen den Untersuchungsgruppen gibt. Die Interventionsgruppe zeigt über alle Messzeitpunkte hinweg geringere Werte in der Fremdgruppenabwertung gegenüber einem chinesischen Kind als die Kontrollgruppe. Zum dritten Messzeitpunkt liegen die Werte der Interventionsgruppe sogar im negativen Bereich, was einer minimalen Eigengruppenabwertung gleichkommt.

Das bedeutet, dass die von Anfang an gering ausgeprägte kognitive Fremdgruppenabwertung gegenüber einem chinesischen Kind in beiden Untersuchungsgruppen über die Zeit weiter abnimmt. Dabei ist in der Gesamtgruppe kein Interventionseffekt auszumachen, während in der Risikogruppe 25 zumindest zwischen dem zweiten und dritten Messzeitpunkt eine Tendenz in die erwartete Richtung deutlich wird (s. Anhang D).

Diese Befunde werden durch die Ergebnisse der Verkuyten-Skala gegenüber der Gruppe chinesischer Kinder gestützt. Auch hier findet sich zum ersten Messzeitpunkt nur eine leichte Eigengruppenbevorzugung im Vergleich zur chinesischen Fremdgruppe. Dabei liegen hier allerdings die Prätestwerte der Interventionsgruppe sogar signifikant unter denen der Kontrollgruppe. In beiden Untersuchungsgruppen findet ein Abfall der Werte zum Posttest hin statt. Hier zeigt sich auch ein signifikanter Unterschied zwischen den Untersuchungsgruppen, der nicht allein auf den Prätest oder die Interaktion von Prätest und Gruppenzugehörigkeit zurückzuführen ist. Die Kontrollgruppe hat minimal mehr an Einstellungsverzerrungen verloren, als die Interventionsgruppe. Dieser Effekt ist jedoch so klein, dass er praktisch zu vernachlässigen ist, wie sich an der nahe Null liegenden Prä-Post-Netto-Effektstärke ablesen lässt. Schließlich steigen die Werte in der Interventionsgruppe wieder minimal an, was die sehr kleine positive Netto-Effektstärke im Prä-Follow-up-Vergleich abbildet. Allerdings liegt auch zum zweiten und dritten Messzeitpunkt das Niveau der Einstellungsverzerrungen in der Interventionsgruppe niedriger als in der Kontrollgruppe.

Das bedeutet, dass die zunächst unterschiedlichen Niveaus der beiden Untersuchungsgruppen in der kognitiven Eigengruppenbevorzugung gegenüber der Gruppe chinesischer Kinder sich über die Zeit einander annähern.

Zusammenfassung

Insgesamt kann festgehalten werden, dass es signifikante Unterschiede zwischen den Untersuchungsgruppen bei den gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt gibt. Die Interventionsgruppe liegt fast zu allen Messzeitpunkten auf einem leicht niedrigeren Niveau der Einstellungsverzerrungen gegenüber der chinesischen Fremdgruppe als die Kontrollgruppe. Weiterhin zeigen sich unterschiedliche Entwicklungsverläufe insbesondere gegenüber der russischen Fremdgruppe und hier besonders zwischen dem zweiten und dritten Messzeitpunkt. Hier findet sich in der Interventionsgruppe ein Absinken der Einstellungsverzerrungen, während es in der Kontrollgruppe stabile Werte oder einen leichten Wiederanstieg zu verzeichnen gibt. Dies gilt sowohl für die

gruppenzentrierte Verkuyten-Skala²⁸, als auch für die personenzentrierten MRA-Skalen.

Einordnung der Ergebnisse zu gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt

Im Vergleich zu den Ergebnissen anderer Evaluationsstudien ergibt sich für die hier vorliegenden Ergebnisse zu gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt ein differenziertes Bild. Betrachtet man die Ergebnisse der Stellvertretender-Kontakt-Studien der Arbeitsgruppe um Lindsey Cameron (2006a, 2006b, 2007) zeigen sich dort signifikante Effekte der Intervention mit mittleren bis großen Effektstärken²⁹ auf den Fremdgruppeneinstellungen der Zielfremdgruppen des verwendeten MRA. Im Vergleich dazu finden sich bei PARTS auf den Biaswerten der Einstellungsskalen mit kognitivem Schwerpunkt nur kleine und verzögerte Interventionseffekte. Allerdings können diese Studien nicht so einfach mit der PARTS-Studie verglichen werden, denn hier zeigt sich bereits der erste Unterschied zur vorliegenden Untersuchung. Es wurden bei Cameron und Kollegen reine Fremdgruppeneinstellungen und Eigengruppeneinstellungen berechnet und keine Bias-Werte, wie in der vorliegenden Arbeit. Dieses Vorgehen (von Cameron und Kollegen) lässt jedoch generelle Antworttendenzen der einzelnen Kinder unberücksichtigt in die Ergebnisse einfließen, während diese bei der Betrachtung von Bias-Werten herausgerechnet werden. Des Weiteren zeigen die Studien von Cameron und Kollegen Mängel im Untersuchungsdesign, da entweder auf eine Kontrollgruppe oder den Prätest verzichtet wurde. Somit ist eine Berechnung von Netto-Effektstärken dort nicht möglich und Validitätsgefährdungen der Ergebnisse können nicht ausgeschlossen werden. Weiterhin sind die gefundenen Interventionseffekte der britischen Studien nur kurzfristig abgesichert. Der Posttest fand eine Woche nach Ende der Intervention statt und es gab keine Follow-up-Untersuchungen.

Bei den aktuelleren wissensbasierten Evaluationsstudien (Perkins & Mebert, 2005; Whetstone, 1992) zeigen sich dagegen, trotz vergleichbarer

²⁸ Soweit man das auf Basis der Datenlagen der kognitiven Verkuyten-Skala gegenüber der Gruppe russischer Kinder feststellen kann, da hier ausschließlich Daten zum dritten Messzeitpunkt vorliegen. Diese weisen aber auf eine entsprechend äquivalente Entwicklung hin.

²⁹ Effektstärken wurden nachträglich vom Autor berechnet und basieren, bis auf Cameron und Rutland (2006), auf Interventions-Kontrollgruppen-Vergleichen innerhalb des Posttest.

Gesamtinterventionsintensität, keine ethnien-spezifischen Effekte auf dem auch hier verwendeten MRA. In Übersichtsarbeiten (Stephan, Renfro & Stephan, 2004; McGregor, 1993) finden sich jedoch kleine bis mittlere kurzfristige und große längerfristige Effekte, allerdings bei Vorurteilen allgemein und nicht speziell bei gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt.

Bei Evaluationsstudien mit sozial-kognitiven Interventionen mit einer ähnlichen oder höheren Interventionsintensität als PARTS, zeigen sich kurzfristig mittlere bis große Interventionseffekte auf kognitiven Einstellungsmaßen (Bigler & Liben, 1992; Weiner & Wright, 1973; Aboud & Fenwick, 1999). Bei der zweiten von Aboud und Fenwick beschriebenen Studie (Aboud & Doyle, 1996) sind Interventionsinhalte und Effektmaße allerdings fast identisch, so dass hier trotz der sehr kurzen Interventionsdauer ein positives Ergebnis zu erwarten war. Außerdem zeigen sich bei den Studien von Aboud die Effekte nur bei der im Prätest identifizierten Gruppe von Teilnehmern mit hohen Vorurteilen. Somit könnten hier, zumindest teilweise, ebensogut Regressionseffekte zur Mitte für die angeblichen Interventionseffekte verantwortlich sein.

Insgesamt erweisen sich die Interventionseffekte auf gruppenbezogene soziale Einstellungen mit kognitivem Schwerpunkt bei PARTS im Vergleich zu den Ergebnissen aus der Literatur als kleiner. Dies könnte mit der zum Teil geringeren Intensität der einzelnen Interventionsbereiche von PARTS im Vergleich zu anderen Interventionen zusammenhängen (s.o. Intensität und Dauer). Außerdem zeigen sich bei den kognitiven gruppenbezogenen sozialen Einstellungen die Effekte bei PARTS erst zum Follow-up, während es in der Literatur hauptsächlich kurzfristige Interventionseffekte gibt und kaum Studien mit einer längerfristigen Prüfung von Interventionseffekten. Auf diese verzögerten sogenannten Sleeper-Effekte von PARTS wird in der generellen Diskussion näher eingegangen. Schließlich finden sich die Interventionseffekte bei PARTS jedoch auf verschiedenen Effektmaßen, was für ein valides Ergebnis spricht, allerdings nur bei der russischen Zielfremdgruppe. Die angestrebte Generalisierung der Interventionseffekte auf eine zweite chinesische Fremdgruppe war hier nicht erfolgreich. Es zeigen sich eher gegenteilige Entwicklungstendenzen.

Gruppenbezogene soziale Einstellungen mit affektivem Schwerpunkt

Die gruppenbezogenen sozialen Einstellungen mit affektivem Schwerpunkt³⁰ liegen, ebenso wie die gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt, insgesamt auf einem sehr niedrigen Level. Dabei fällt auch bei diesen Einstellungsmaßen ein leichter genereller Niveauunterschied zwischen der chinesischen und der russischen Fremdgruppe auf, wobei die Einstellungsverzerrungen gegenüber der russischen Fremdgruppe stets etwas stärker ausgeprägt sind, als gegenüber der chinesischen Fremdgruppe. Nichts desto trotz liegen auch die Einstellungsverzerrungen gegenüber der russischen Gruppe immer noch auf einem sehr niedrigen Niveau.

Affektive Einstellungen gegenüber der russischen Zielfremdgruppe

Die IBM-Sympathie-Einschätzungen gegenüber einem russischen Kind zeigen zum dritten Messzeitpunkt, ähnlich wie bei der Verkuyten-Sympathie-Einschätzung (s.u.), einen signifikanten Niveauunterschied zwischen den Untersuchungsgruppen. Die Teilnehmer der Interventionsgruppe zeigen weniger affektive Einstellungsverzerrungen gegenüber der russischen Fremdgruppe, als die Teilnehmer der Kontrollgruppe. Dies drückt sich auch in der kleinen negativen Effektstärke innerhalb des dritten Messzeitpunktes aus.

Ein ähnlicher Befund findet sich bei der Verkuyten-Sympathie-Einschätzung. Die Interventionsgruppe zeigt hier weniger Eigengruppenbevorzugung gegenüber der Gruppe russischer Kinder. Dies macht sich ebenfalls durch eine kleine negative Effektstärke bemerkbar.

Das bedeutet, dass die Kinder der Interventionsgruppe zum dritten Messzeitpunkt signifikant weniger affektive Einstellungsverzerrungen gegenüber der Gruppe russischer Kinder zeigen als die Teilnehmer der Kontrollgruppe.

Auch im Bereich der sozialen Distanz zeigt die Interventionsgruppe eine geringere Einstellungsverzerrung gegenüber der russischen Zielfremdgruppe. Dabei finden sich auch hier signifikante Mittelwertsunterschiede zwischen den beiden Untersuchungsgruppen zu t2 und zu t3. Die sich ergebenden negativen Effektstärken liegen hier zwischen dem kleinen und dem mittleren Bereich.

³⁰ Dazu zählen die IBM-Sympathie-Einschätzung, die Verkuyten-Sympathie-Einschätzung, sowie die soziale Distanz-Skala „Haus-Aufgabe“.

Das bedeutet, dass die Interventionsgruppe sowohl zum Posttest als auch zum Follow-up signifikant weniger affektive Einstellungsverzerrungen in Form von sozialer Distanz gegenüber einer russischen Familie zeigt, als die Kontrollgruppe.

Insgesamt zeigt die Interventionsgruppe zu t2 und t3 signifikant weniger affektive Einstellungsverzerrungen gegenüber der russischen Zielfremdgruppe als die Kontrollgruppe. Die Effektstärken liegen dabei im kleinen bis etwa mittleren Größenbereich.

Affektive Einstellungen gegenüber r der chinesischen Fremdgruppe

Die Werte der IBM-Sympathie-Einschätzungen liegen auf einem sehr niedrigen Level. Dabei steigen die IBM-Sympathie-Einschätzungen gegenüber einem chinesischen Kind über die Zeit in beiden Gruppen leicht an. Hier nimmt zunächst die Interventionsgruppe stärker zu als die Kontrollgruppe, was sich zum Follow-up hin allerdings umkehrt. Da bleibt die Interventionsgruppe in ihren Werten stabil, während die Kontrollgruppe weiter zunimmt. Die Interventionsgruppe zeigt also verzögert einen geringeren Anstieg als die Kontrollgruppe. Diese unterschiedlichen Entwicklungsverläufe werden durch eine kleine positive Prä-Post-Netto-Effektstärke und eine kleine negative Prä-Follow-up-Netto-Effektstärke verdeutlicht und auch statistisch signifikant. Nur der post-hoc-Kontrast zwischen dem zweiten und dem dritten Messzeitpunkt erreicht dabei das Signifikanzniveau.

Das bedeutet, dass die affektiven Einstellungsverzerrungen gegenüber einem chinesischen Kind in beiden Untersuchungsgruppen ansteigen. Dieser Anstieg fällt allerdings zwischen Posttest und Follow-up in der Interventionsgruppe etwas geringer aus, als in der Kontrollgruppe.

Bei der Verkuyten-Sympathie-Einschätzung ergibt sich ein etwas anderes Bild. Die Prätestwerte der Interventionsgruppe sind hier signifikant niedriger als die Prätestwerte der Kontrollgruppe, bei einem sehr niedrigen Gesamtniveau. Im weiteren Verlauf entwickeln sich beide Gruppen jedoch sehr ähnlich. Zunächst (prä-post) gibt es einen leichten Abfall in den Einstellungsverzerrungen und daraufhin (post-follow-up) einen leichten Wiederanstieg. Dabei lassen sich zwar inferenzstatistisch signifikante Unterschiede zwischen den Gruppen finden, allerdings kann dieser Unterschied in der Regression auf den Posttest nicht interpretiert werden, da hier auch der Interaktionsterm der Regressoren Prätest und Gruppe bedeutsam wird. In der Regression auf das Follow-up der affektiven

Einstellungsverzerrungen findet sich jedoch ein signifikanter Regressor Gruppe, ohne durch den Moderator des Interaktionsterms von Prätest und Gruppe überdeckt zu werden. Allerdings spiegelt sich dieser signifikante Unterschied zwischen den Untersuchungsgruppen kaum in den Netto-Effektstärken wieder. Diese liegen nahe Null.

Das bedeutet, dass die affektive Eigengruppenbevorzugung gegenüber der Gruppe chinesischer Kinder zwischen dem ersten und dem zweiten Messzeitpunkt in beiden Untersuchungsgruppen leicht abfällt. Im weiteren Verlauf stabilisieren sich diese Einstellungen auf dem Niveau des Posttests. Hierbei zeigen sich keine Interventionseffekte.

Die kleinen Niveau-Unterschiede zwischen den beiden Untersuchungsgruppen bei den affektiven Einstellungsverzerrungen gegenüber der chinesischen Fremdgruppe zeigen sich auch auf der sozialen Distanz-Skala gegenüber einer chinesischen Familie. Auch hier zeigen sich signifikante Mittelwertsunterschiede zwischen den Untersuchungsgruppen. Beim Posttest zeigt die Interventionsgruppe weniger affektive Eigengruppenbevorzugung als die Kontrollgruppe. Dies wird auch an der kleinen negativen Effektstärke innerhalb des zweiten Messzeitpunktes deutlich. Die Werte bleiben bis zum dritten Messzeitpunkt in beiden Gruppen relativ stabil, und zeigen auch dort einen signifikanten Mittelwertsunterschied und eine kleine negative Effektstärke.

Das bedeutet, dass die Teilnehmer der Interventionsgruppe sowohl zum zweiten als auch zum dritten Messzeitpunkt signifikant weniger affektive Einstellungsverzerrungen in Form von sozialer Distanz gegenüber einer chinesischen Familie zeigen, als die Teilnehmer der Kontrollgruppe.

Zusammenfassung

Insgesamt lässt sich in diesem Einstellungsbereich ein positives Bild der Interventionseffekte, vor allem gegenüber der russischen Fremdgruppe, zeichnen.

Es gibt signifikante Unterschiede zwischen den Untersuchungsgruppen auf den gruppenbezogenen sozialen Einstellungen mit affektivem Schwerpunkt. Diese zeigen sich allerdings hauptsächlich gegenüber der russischen Fremdgruppe zum zweiten und dritten Messzeitpunkt. Diese Effekte sind zwar klein, aber statistisch bedeutsam und konsistent über verschiedene Maße hinweg auffindbar.

Einordnung der Ergebnisse zu gruppenbezogenen sozialen Einstellungen mit affektivem Schwerpunkt

Bei der Einordnung der oben zusammengefassten Ergebnisse zu gruppenbezogenen sozialen Einstellungen mit affektivem Schwerpunkt in den literarischen Kontext zeigt sich zunächst ein großer Mangel in der Anwendung solcher Maße bei anderen Evaluationsstudien. Bei den Stellvertretender-Kontakt-Studien der Arbeitsgruppe um Lindsey Cameron wird auf affektive Einstellungsmaße verzichtet. Bei den meisten wissenschaftlichen Interventionsstudien sieht es ähnlich aus. Eine Ausnahme bildet hier die Studie von Lauren Whetstone (1992). Allerdings konnte sie keine Interventionseffekte auf der von ihr verwendeten Skala zur sozialen Distanz gegenüber Fremdgruppen finden. Ansonsten bleibt für die Einordnung der Ergebnisse wissenschaftlicher Interventionen hier nur der Rückgriff auf die bereits oben erwähnten positiven Ergebnisse der Metaanalysen aus diesem Bereich (Stephan, Renfro & Stephan, 2004; McGregor, 1993), auch wenn abermals angemerkt werden muss, dass die dort beschriebenen kleinen bis mittleren kurzfristigen bzw. großen langfristigen Effekte dem allgemeinen Vorurteilsbegriff, und nicht speziell affektiven Einstellungsmaßen, zugeordnet werden müssen.

In zwei älteren Studien von Katz (1973) sowie Katz und Zalk (1978) finden sich jedoch Ergebnisse sozial-kognitiver Interventionen auf einer sozialen Distanz-Skala. Hier konnte zumindest kurzfristig ein signifikanter positiver Effekt der Intervention zur differenzierten Wahrnehmung von Fremdgruppenmitgliedern bei der sozialen Distanz gefunden werden. Leider wurden hier weder Effektstärken berechnet, noch ließen sich auf Grund mangelnder Angaben in den Artikeln nachträglich solche berechnen. Der positive Effekt fand sich allerdings auch nur unter bestimmten Bedingungen (Alter der Kinder und Ethnie des Experimentators) und verschwand zum Follow-up nach vier Monaten wieder, was ein Hinweis auf einen eher kleinen Effekt sein kann (Katz & Zalk, 1978). Abgesehen davon, dass auch hier nur Teilnehmer in der Interventionsgruppe waren, die zum Prätest erhöhte Vorurteilswerte gezeigt haben.

Somit finden sich in diesem Bereich nur wenige Befunde in der Literatur, die zur Einordnung der Ergebnisse der vorliegenden Arbeit nützen könnten.

Die hier vorliegenden Effekte weisen jedoch darauf hin, dass PARTS durchaus eine Wirkung auf affektive Einstellungsmaße der Teilnehmer hat; dies besonders bei der russischen Zielfremdgruppe. Die kleinen aber signifikanten

Unterschiede finden sich auf drei verschiedenen Einstellungsmaßen aus diesem Bereich, was für einen validen Effekt spricht, auch wenn zu all diesen Maßen leider keine Prätestwerte für die russische Fremdgruppe vorliegen. Die Alternativerklärung, dass bereits zum Prätest derartige Niveau-Unterschiede zwischen den beiden Untersuchungsgruppen auf drei verschiedenen Messinstrumenten bestehen und bis zum dritten Messzeitpunkt erhalten bleiben, scheint äußerst unwahrscheinlich. Zumal sich abgesehen von den Verkuyten-Skalen gegenüber der chinesischen Fremdgruppe in keinem Bereich signifikante Gruppenunterschiede zum Prätest zeigen.

Dieses Ergebnis passt weiterhin zu der Argumentation von Dovidio und Kollegen (2004), die bei Kontakt-Interventionen eher affektive Wirkprozesse vermuten und dementsprechend auch eine verstärkte Wirkung auf affektive Effektmaße gegenüber der russischen Zielfremdgruppe aus der Stellvertretender-Kontakt-Intervention erwarten lassen.

Die Ergebnisse zur chinesischen Fremdgruppe sind nicht so einheitlich. Hier finden sich nur auf der Hausaufgabe sehr kleine signifikante Interventionseffekte sowie ein ebenfalls sehr kleiner signifikanter Effekt auf dem affektiven IBM-Item zum dritten Messzeitpunkt. Das affektive Verkuyten-Item hingegen weist eher in die gegensätzliche Richtung und spricht für einen negativen Effekt der Intervention auf die affektive Eigengruppenbevorzugung gegenüber der Gruppe chinesischer Kinder. Diese Effekte sind jedoch so klein, dass sie praktisch nicht bedeutsam sind. Sie basieren vermutlich auf dem vorliegenden signifikanten Unterschied im Prätest zwischen den Untersuchungsgruppen. Auch im Bereich der affektiven Einstellungsmaße scheint eine Generalisierung der Interventionseffekte von der Zielfremdgruppe auf andere Fremdgruppen nicht gelingen zu sein.

Gruppenbezogene soziale Einstellungen mit konativem Schwerpunkt

Die Werte der gruppenbezogenen sozialen Einstellungen mit konativem Schwerpunkt³¹ bewegen sich, ähnlich wie in den anderen beiden gruppenbezogenen Einstellungsbereichen, insgesamt auf einem sehr niedrigen Niveau.

Konative Einstellungen gegenüber der russischen Zielfremdgruppe

Die gruppenbezogenen sozialen Einstellungen mit konativem Schwerpunkt gegenüber einem russischen Kind zeigen die auch in anderen Einstellungsbereichen gefundenen Mittelwertsunterschiede zwischen der Interventionsgruppe und der Kontrollgruppe zum dritten Messzeitpunkt. Die Interventionsgruppe zeigt hier signifikant weniger Eigengruppenbevorzugung auf dem IBM-Maß gegenüber dem russischen Kind, als die Kontrollgruppe. Die Effektstärke ist zwar klein, aber vorhanden.

Dies bedeutet, dass die Teilnehmer der Interventionsgruppe zum dritten Messzeitpunkt signifikant eher bereit sind, mit einem russischen Kind Kontakt aufzunehmen als die Teilnehmer der Kontrollgruppe.

Konative Einstellungen gegenüber der chinesischen Fremdgruppe

Hier findet über die Zeit ein leichter Anstieg in den Einstellungsverzerrungen gegenüber der chinesischen Fremdgruppe statt. Allerdings steigert sich zwischen Prätest und Posttest zunächst die Interventionsgruppe etwas mehr als die Kontrollgruppe und zwischen Posttest und Follow-up die Kontrollgruppe mehr als die Interventionsgruppe. Dies führt zu den sehr kleinen, einmal positiven, einmal negativen Netto-Effektstärken. Diese unterschiedlichen Entwicklungsverläufe erreichen jedoch nicht das Signifikanzniveau.

Dies bedeutet, dass die Einstellungsverzerrungen mit konativem Schwerpunkt gegenüber einem chinesischen Kind über die Zeit leicht zunehmen bzw. die Kontaktbereitschaft zu einem chinesischen Kind leicht abnimmt. Dabei bleibt jedoch die Interventionsgruppe tendentiell auf einem niedrigeren Niveau als die Kontrollgruppe.

³¹ Dazu zählt nur der IBM.

Zusammenfassung

Insgesamt zeigt sich ein kleiner, aber signifikanter Interventionseffekt auf der konativen Einstellungsverzerrung gegenüber einem russischen Kind, nicht jedoch gegenüber Kindern der chinesischen Fremdgruppe.

Einordnung der Ergebnisse zu gruppenbezogenen sozialen Einstellungen mit konativem Schwerpunkt

In der Literatur zeigen sich in diesem Einstellungsbereich unterschiedliche Ergebnisse. Bei den Stellvertretender-Kontakt-Studien (Cameron et al., 2006a, 2006b, 2007) finden sich teilweise große signifikante Effekte (Cameron und Rutland, 2006) auf den Kontaktintentionen des auch hier verwendeten Intended Behavior Measure. In zwei anderen Studie zeigen sich allerdings keine Auswirkungen der SK-Intervention auf die Verhaltensintentionen (Cameron et al., 2006b, 2007). Doch unter der Bedingung, hoch mit der eigenen Nationalität identifiziert zu sein, fand sich auch in der letzten Studie von Cameron und Kollegen (2007) noch ein großer signifikanter Effekt auf den gemessenen Verhaltensintentionen gegenüber der Flüchtlingszielfremdgruppe. Hier ergibt sich also ein uneinheitliches Bild. Dabei gelten für diese Ergebnisse die gleichen Einschränkungen, die bereits oben zu der Einordnung der gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt genannt wurden. Lücken im Untersuchungsdesign gefährden die interne Validität, nur kurzfristige Posttestmessungen sichern die Effekte nicht längerfristig ab. Es wurden keine Bias-Werte sondern singuläre Einstellungswerte betrachtet, womit Antworttendenzen der Kinder vernachlässigt wurden.

Zu den wissensbasierten Interventionen finden die Autoren der Metaanalyse (Stephan et al., 2004), wie auch bei Vorurteilen allgemein, kurzfristig kleine und langfristig große Interventionseffekte auf Verhaltensmaßen. Dabei muss jedoch beachtet werden, dass die langfristigen Effekte nur auf wenigen, vermutlich besonders hochwertigen Studien beruhen. Hier könnte also auch eine Art von Selektionseffekt vorliegen, der zur Überschätzung der Effektgrößen führen kann.

Auch in einigen wenigen Studien zu sozial-kognitiven Interventionen ließen sich bedeutsame Interventionseffekte auf Verhaltensintentionen (Weiner & Wright, 1973), oder gar auf direkten Verhaltensmaßen zur sozialen Distanz im Raum und zum Imitationsverhalten von gegenethnischen Versuchsleitern finden (Katz und Zalk, 1978). Diese Effekte konnten zum Teil (Katz & Zalk, 1978) sogar in einem

viermonatigen Follow-up noch bestätigt werden. Allerdings muss hierzu angemerkt werden, dass die Interventionsgruppe bei Katz und Zalk aus hoch vorurteilsbelasteten Kindern bestand und somit die Regression zur Mitte möglicherweise diesen Effekt mit gestützt hat. Leider ließen sich bei diesen Studien keine Effektstärken berechnen, womit die Größe der Effekte fraglich bleibt.

Insgesamt zeigt sich hier ähnlich wie bei den Ergebnissen zu gruppenbezogenen sozialen Einstellungen mit kognitivem Schwerpunkt, dass die Interventionseffekte von PARTS etwas geringer sind als die Effekte, die sich zumindest teilweise in der Literatur finden lassen. Dies lässt sich im Falle der Stellvertretender-Kontakt-Studien mit der geringeren Intensität der SK-Intervention bei PARTS erklären.

Leider verfügt die vorliegende Studie auch bei dieser Skala gegenüber der russischen Zielfremdgruppe nur Daten zum dritten Messzeitpunkt. Daher kann nicht beurteilt werden, ob auch bei den konativen Einstellungsmaßen ähnlich wie bei den kognitiven Einstellungsmaßen der Effekt verzögert eintritt. Die entsprechenden Ergebnisse zur chinesischen Fremdgruppe, bei denen erst zwischen t2 und t3 ein tendentieller Interventionseffekt deutlich wird, können aber als Hinweis für eine ähnliche Entwicklung in diesem Bereich gewertet werden.

Leider ist es auch bei den konativen Einstellungsmaßen nicht gelungen die kleinen Interventionseffekte der russischen Zielfremdgruppe auf die chinesische Fremdgruppe zu generalisieren.

Interpersonale soziale Einstellungen und allgemeines Sozialverhalten

Insgesamt zeigt sich auch in den interpersonalen sozialen Einstellungen und dem generellen Sozialverhalten ein sehr gutes Funktionsniveau der Kinder. Durchschnittlich erzielen sie hohe Ergebnisse bei positiven Einstellungen und prosozialen Verhaltenseinschätzungen und niedrige Werte bei den negativen Einstellungen und den problematischen Verhaltenseinschätzungen.

Interpersonale soziale Einstellungen: Skala Akzeptanz

Die Werte bei der interpersonalen Einstellungsskala Akzeptanz liegen bereits zum ersten Messzeitpunkt auf einem sehr hohen Niveau. Es findet sich zudem ein minimaler Anstieg über die Zeit in beiden Untersuchungsgruppen, aber keine bedeutsamen Gruppenunterschiede. Dies zeigt sich auch an Hand der nahe Null liegenden positiven Netto-Effektstärken.

Das bedeutet, dass die Akzeptanz der Kinder in beiden Gruppen über die Zeit leicht ansteigt. Nur in der Risikogruppe 25 findet sich ein tendentieller kleiner Beschleunigungseffekt zwischen dem ersten und dem zweiten Messzeitpunkt in der Interventionsgruppe (s. Anhang D).

Interpersonale soziale Einstellungen: Skala Intoleranz

Im Bereich Intoleranz hingegen liegen die Werte in beiden Untersuchungsgruppen zunächst auf einem mittleren Skalenniveau. Hier zeigt sich trotz des gemeinsamen Abfalls der Intoleranzwerte ein kurzfristiger signifikanter Interventionseffekt. Die Werte der Interventionsgruppe zwischen t1 und t2 sinken stärker ab, als in der Kontrollgruppe. Dies zeigt sich auch an der kleinen negativen Prä-Post-Netto-Effektstärke. Dieser Unterschied nivelliert sich jedoch zum Follow-up wieder.

Das bedeutet, dass die Intoleranz der Teilnehmer an der Untersuchung über die Zeit abnimmt. Dabei sinkt die Intoleranz in der Interventionsgruppe zunächst signifikant stärker, als die in der Kontrollgruppe. Langfristig gleichen sich die Einstellungsniveaus zwischen den Gruppen jedoch wieder an. Ein ähnlicher Beschleunigungseffekt findet sich auch bei der Analyse der Risikogruppe 25 (s. Anhang D).

Einordnung der Ergebnisse zu interpersonalen Einstellungen

Will man die Ergebnisse zu den interpersonalen Einstellungen mit Ergebnissen aus der Literatur vergleichen, wird man enttäuscht. Es finden sich keine Studien zur Prävention von Vorurteilen, die auch positive Einstellungen wie Toleranz messen. Daher war es auch notwendig, einen neuen Fragebogen zur interpersonalen Toleranz selbst zu entwickeln. Dies ist, vor allem in Bezug auf die Subskala der Intoleranz, auch gut gelungen (s. Kapitel III.6. Messinstrumente).

Was die Befunde der vorliegenden Arbeit auf den beiden interpersonalen Einstellungsskalen angeht, scheint hier durch die Intervention ein Beschleunigungseffekt in der Interventionsgruppe vorzuliegen. Auf der Intoleranzskala zeigt sich der Effekt signifikant mit kleiner Effektstärke, während sich auf der Akzeptanzskala nur eine leichte Tendenz in diese Richtung abzeichnet, besonders wenn man hier die Ergebnisse der Risikogruppe 25 miteinbezieht. Dieser Beschleunigungseffekt könnte bei der Akzeptanzskala durch einen Deckeneffekt auf der Skala ausgebremst werden und kommt daher eventuell nicht voll zur Geltung. Ansonsten verwundert es nicht, dass die Interventionsgruppe hier einen zumindest kurzfristigen Vorteil gegenüber der Kontrollgruppe erlangt. Schließlich wurden in der Interventionsgruppe Übungen zu Themen wie Perspektivenübernahme, Anerkennung anderer Meinungen, soziales Problemlösen etc. durchgeführt. Augenscheinlich ergibt sich eine derartige Weiterentwicklung der Kinder in diesem Alter allerdings auch ohne Intervention mit zunehmender Reife.

Sozialverhalten: Gesamtproblemwert des Strengths and Difficulties Questionnaire

Beim Sozialverhalten, eingeschätzt durch die Eltern der Teilnehmer, liegen die Werte auf einem sehr niedrigen Niveau. Die Werte bleiben in beiden Gruppen über die Zeit relativ stabil. Nur in der Interventionsgruppe findet sich zwischen zweitem und drittem Messzeitpunkt ein minimaler nicht signifikanter Abfall der Werte. Dies zeigt sich auch in der sehr kleinen negativen Prä-Follow-up-Netto-Effektstärke.

Das bedeutet, dass die Eltern das Sozialverhalten ihrer Kinder als relativ unproblematisch einschätzen. Diese Einschätzung bleibt stabil, bis auf einen kleinen Abfall des Gesamtproblemwerts in der Interventionsgruppe zwischen dem zweiten und dem dritten Messzeitpunkt, der aber nicht statistisch bedeutsam wird.

Die Werte der Gesamtproblemwert-Skala liegen auch bei den Lehrereinschätzungen auf einem recht niedrigen Level. Hier zeigt sich beim Gesamtproblemwert ein äquivalenter minimaler Abfall in beiden Untersuchungsgruppen vom ersten zum zweiten Messzeitpunkt.

Das bedeutet, dass auch die Lehrer das Sozialverhalten ihrer Kinder als relativ unproblematisch einschätzen. Es lässt sich dabei kein Interventionseffekt feststellen.

Sozialverhalten: Prosoziales Verhalten des Strengths and Difficulties

Questionnaire

Die Elterneinschätzungen des prosozialen Verhaltens ihrer Kinder bleiben über die Zeit in beiden Untersuchungsgruppen, ähnlich wie die Werte der Gesamtproblemskala, recht stabil. Es findet nur ein minimaler Anstieg der Werte über die Zeit statt. Dieser Anstieg ist in der Kontrollgruppe leicht größer, als in der Interventionsgruppe, was die winzigen statistisch nicht bedeutsamen negativen Netto-Effektstärken erzeugt.

Dies bedeutet, dass das prosoziale Verhalten der Kinder zu Hause über die Zeit relativ stabil bleibt. Es finden sich dabei kaum Unterschiede zwischen den Untersuchungsgruppen. Einzig ein Trend für einen Beschleunigungseffekt in der Interventionsgruppe kann bei der Risikogruppe 25 verzeichnet werden (s. Anhang D).

Auch bei den Lehrereinschätzungen liegen die Werte des prosozialen Verhaltens auf einem hohen Niveau. Hier findet sich außerdem ein leichter Anstieg der Werte der Interventionsgruppe, während die Werte der Kontrollgruppe leicht absinken. Diese signifikant unterschiedlichen Entwicklungsverläufe der beiden Untersuchungsgruppen führen zu einer kleinen positiven Netto-Effektstärke im Prä-Post-Vergleich.

Dies bedeutet, dass sich das prosoziale Verhalten der Teilnehmer der Interventionsgruppe im Schulkontext im Vergleich zum prosozialen Verhalten in der Kontrollgruppe signifikant steigert. Allerdings ist auch hier die positive Prä-Post-Netto-Effektstärke nur als klein einzustufen.

Einordnung der Ergebnisse zum allgemeinen Sozialverhalten

Zur Einordnung der Ergebnisse zum Sozialverhalten der Kinder kann auf die Studien von Johnson und Kollegen (1994, 1995) verwiesen werden. Hier gelang es

den Autoren durch die Vermittlung einer Verhandlungs- und Mediationsstrategie, das Sozialverhalten der Kinder nachhaltig zu beeinflussen. Dies zeigte sich zum einen an den verwendeten Konfliktszenarien zum anderen aber auch an der geringeren Anzahl von Konflikten zwischen den Kindern, die vom Lehrpersonal in der Folge des Trainings bearbeitet werden mussten. Hier wurde das Sozialverhalten der Teilnehmer durch ein sehr intensives lang andauerndes Training, 9 bis 15 Stunden über 6 Wochen verteilt, insgesamt positiv beeinflusst.

Im Vergleich dazu wurde bei PARTS lediglich 70 Minuten über 7 Wochen das soziale Problemlösen an Hand von Szenarien mit den Kindern geübt. Hier ist also eine deutlich geringere Intensität dieses Trainingsbereichs zu verzeichnen. Dies erklärt vermutlich auch, warum sich nur kurzfristige Effekte (prosoziales Verhalten der Lehrereinschätzung) bzw. tendentielle Beschleunigungseffekte (prosoziales Verhalten Elterneinschätzung in der Risikogruppe 25) in der vorliegenden Arbeit zum Sozialverhalten finden lassen. Bezeichnend ist auch, dass die gefundenen Effekte hier eher im Kompetenzbereich liegen und nicht in der Abnahme von Problemverhalten. Dies ist ein Hinweis für den Erfolg der Förderung positiver Einstellungen wie Toleranz und Akzeptanz sowie des allgemeinen Sozialverhaltens. Ob sich der Effekt auf dem prosozialen Verhalten der Lehrereinschätzung auch längerfristig zeigt, kann auf Grund der an dieser Stelle mangelnden Datenlage leider nicht beurteilt werden³². Interessant ist bei diesem Befund auch, dass sich der Interventionseffekt auf das prosoziale Verhalten der Kinder nur im Schulkontext belegen lässt. Dies könnte ein Hinweis dafür sein, dass PARTS nur kontextspezifisch wirksam ist, und eine Generalisierung der Effekte schwierig ist.

Insgesamt zeigen sich bei den interpersonalen sozialen Einstellungen und beim Sozialverhalten zumindest kurzfristig positive Effekte der Intervention. Die Intoleranz nimmt stärker ab als in der Kontrollgruppe, das prosoziale Verhalten steigt, zumindest im Schulkontext, stärker an als in der Kontrollgruppe. Hier scheinen Beschleunigungseffekte der Intervention vorzuliegen.

³² Es wurde davon abgesehen, die Lehrer auch zum dritten Messzeitpunkt zu befragen, um die ohnehin hohe Belastung des Lehrpersonals nicht weiter zu strapazieren und eine positive Zusammenarbeit mit dem Forscherteam auch weiterhin zu gewährleisten.

V.6. *Kernpunkte der Befunde, Erläuterungen und Zielerreichung von PARTS*

Im Folgenden werden die Kernpunkte der oben geschilderten Befunde zur Implementation und Wirksamkeit von PARTS überblicksartig zusammengefasst. Dazu werden die Befunde mit den Zielen von PARTS abgeglichen.

Strukturelle Parameter von PARTS

PARTS ist ein wissenschaftlich fundiertes multimodales Präventionsprogramm zur Verringerung negativer gruppenbezogener sozialer Einstellungen. Es besteht aus den drei Interventionsbereichen Förderung sozial-kognitiver Fähigkeiten, wissensbasierte Intervention und Stellvertretender-Kontakt, und es verfügt dabei über einen sozial-kognitiven Fokus.

Die Intensität und Dauer des gesamten Trainingsprogramms PARTS liegt leicht über dem Durchschnitt der in der Literatur genannten Programme. Hier liegen zwei Drittel der Programme im Bereich von 1 bis 10 Sitzungen (Beelmann & Heinemann, in Vorbereitung), während sich PARTS über 14 Sitzungen erstreckt. Allerdings weisen die einzelnen Interventionsbereiche eine geringere Trainingsintensität auf, als dies bei entsprechenden unimodalen Programmen der Fall ist.

Implementation

Die Implementation von PARTS hat insgesamt hervorragend funktioniert. Die Teilnehmer engagierten sich mit großer Freude und Interesse an den Programminhalten, was für einen hohen Aufforderungscharakter der Spiele und Übungen von PARTS spricht. Außerdem zeigten die Kinder sehr wenig Störverhalten, was als Hinweis für entwicklungsangemessene Aufgaben und ein angemessenes Schwierigkeitsniveau der Programminhalte gewertet werden kann. Schließlich zeigten sich diese positiven Implementationsergebnisse relativ unabhängig von Programmbereichen und Personenmerkmalen. Dies spricht für den Einsatz von PARTS als universelles Präventionsprogramm.

Proximale Wirksamkeit

Die proximalen Ziele von PARTS wurden nur teilweise erreicht. So gelang es sehr gut, das interkulturelle Wissen der Teilnehmer kurzfristig und langfristig zu fördern. Hier zeigt sich zusätzlich zu den kurzfristigen großen Effekten noch eine Vergrößerung der Effektstärken zum Follow-up. Das Trainingsprogramm scheint in diesem Bereich eine selbstständige Weiterentwicklung der Teilnehmer angeregt zu haben. Ähnliche, sich steigernde Effekte wurden zumindest im Einstellungsbereich auch in der Literatur vorgefunden (vgl. Stephan et al., 2004).

Bei den sozial-kognitiven Fähigkeiten zeigen sich jedoch, wenn überhaupt, nur sehr kleine tendentielle Effekte des Trainingsprogramms. Dies liegt vermutlich an der mangelnden Intensität der entsprechenden Interventionsanteile, an der zu geringen Abstimmung der Messinstrumente auf die eigentlichen Interventionsinhalte, sowie an teilweise methodischen Problemen wie etwa die sehr geringen internen Konsistenzen der verwendeten Messinstrumente. Zusätzlich muss an dieser Stelle erwähnt werden, dass der Posttest der PARTS-Untersuchung erst zwei Monate nach Beendigung des Trainingsprogramms durchgeführt wurde. Somit könnte ein Teil des Interventionseffekts in diesem Zeitraum schon wieder nachgelassen haben. Im Vergleich dazu wurde in vielen Studien aus der Literatur der Posttest direkt nach der Intervention angesetzt. Dies sollte bei der Beurteilung der kleinen „kurzfristigen“ Interventionseffekte bei PARTS beachtet werden. Unter diesen Umständen können die kleinen positiven Tendenzen auf den sozial-kognitiven Effektmaßen durchaus als Erfolg des Interventionsprogramms gewertet werden.

Der Stellvertretende-Kontakt wurde proximal nicht operationalisiert. Da sich jedoch auf den gruppenbezogenen sozialen Einstellungsmaßen hauptsächlich Interventionseffekte gegenüber der russischen Zielfremdgruppe zeigen, welche ausschließlich bei der Stellvertretender-Kontakt-Intervention Anwendung fand, kann angenommen werden, dass es gelungen ist, einen positiven stellvertretenden-Kontakt zu russischen Kindern herzustellen. Somit scheint dieses proximale Ziel von PARTS gut erreicht worden zu sein.

Distale Wirksamkeit

Zunächst kann an dieser Stelle festgehalten werden, dass eine Unterscheidung der Einstellungsmaße nach personenzentriert versus gruppenzentriert bei dieser Untersuchung eben sowenig notwendig erscheint, wie die

Unterscheidung nach Eigengruppenbevorzugung versus Fremdgruppenabwertung. Die Interventionseffekte finden sich unabhängig von diesen strukturellen Merkmalen auf allen gruppenbezogenen sozialen Einstellungsmaßen. Hier zeigen sich hauptsächlich kleine, meist verzögerte Effekte gegenüber der russischen Zielfremdgruppe. Diese Effekte finden sich sowohl auf kognitiven, als auch affektiven und konativen Einstellungsmaßen. Dabei sind die Effekte jedoch kleiner als die in der hier berichteten Literatur beschriebenen, was wohl der geringeren Intensität der einzelnen Interventionsbereiche geschuldet ist. Ein weiterer Grund für die kleineren Effekte kann in der Gruppengröße bei PARTS liegen. Sowohl die Gesamtstichprobe als auch die Interventionsgruppengröße übersteigt meist die Anzahl der Teilnehmer in anderen Studien. Mit größeren Gruppen ist jedoch nur ein weniger intensives Arbeiten möglich. Dieser Effekt findet sich auch in der Literatur (vgl. Beelmann & Heinemann in Vorbereitung). Des Weiteren liegen die Einstellungswerte der Teilnehmer bereits zu t1 auf einem sehr niedrigen Niveau. Dies lässt wenig Spielraum für ein weiteres Absinken der Bias-Werte – ansonsten würde aus der Eigengruppenbevorzugung eine Fremdgruppenbevorzugung werden – das ist nicht Ziel von PARTS.

Betrachtet man jedoch ein breiteres Spektrum an Interventionsstudien, erweisen sich die Effekte von PARTS, verglichen mit der durchschnittlichen Effektstärke eines Großteils der veröffentlichten Interventionen zur Reduktion von gruppenbezogenen sozialen Einstellungsverzerrungen, die aus der Metaanalyse von Beelmann und Heinemann (in Vorbereitung) hervorgeht ($d = 0.31$), als genau im Durchschnitt liegend. Außerdem sind die Effekte bei PARTS sehr breit gestreut und zeigen sich auf Einstellungsmaßen mit verschiedenen Schwerpunkten. Dies korrespondiert mit der Multimodalität des Interventionsprogramms, das gezielt versucht, verschiedene Einstellungskomponenten anzusprechen. Somit ist der ganzheitliche Präventionsansatz von PARTS, was die Beeinflussung verschiedener Einstellungskomponenten angeht, sehr erfolgreich gewesen.

Dass diese Interventionseffekte verzögert auftreten, kann verschiedene Ursachen haben. Es korrespondiert zum einen mit der Weiterentwicklung im Bereich des interkulturellen Wissens (s.o.), zum anderen in gewissem Maße mit den von Stephan und Kollegen (2004) berichteten Ergebnissen ihrer Metaanalyse zu wissensbasierten Interventionen. Auch hier zeigen sich verzögert größere Interventionseffekte, als in den direkten Postmessungen. Auch hier findet sich also

ein Sleeper-Effekt. Dies könnte bei PARTS der Komplexität des Programms geschuldet sein. Hier könnten sowohl Verarbeitungszeit als auch die Möglichkeit themenbezogene Erfahrungen zu sammeln benötigt werden, um die Wirksamkeit des Programms voll zu entfalten. Pettigrew und Tropp (2008) schlagen zur Klärung der Wirkprozesse bei Kontakt-Interventionen vor, dass hier zunächst affektive Prozesse ablaufen. So soll Kontakt beispielsweise die Intergruppenangst senken. Dies führe zu vermehrtem Kontakt, mehr Wissen und Empathie gegenüber den Fremdgruppenmitgliedern. Somit weite sich der Kontakteffekt dann auch auf andere Einstellungskomponenten aus. Diese These passt zu den Befunden der vorliegenden Arbeit. Hier zeigt sich im Posttest zunächst nur ein Effekt auf der sozialen Distanz gegenüber den Fremdgruppen – einem eher affektiven Einstellungsmaß. Verzögert finden sich dann zum Follow-up Interventionseffekte gegenüber der Zielfremdgruppe auf allen Einstellungsmaßen.

Eine weitere Ursache für die erst verzögert auftretenden Interventionseffekte könnte auch in der normativen Entwicklung der gruppenbezogenen Einstellungen liegen. Denn während die Werte der Interventionsgruppe zwischen dem zweiten und dritten Messzeitpunkt weiter leicht abfallen, oder zumindest stabil bleiben, steigen die Werte in der untrainierten Kontrollgruppe in diesem Zeitraum leicht an. Dieser Anstieg der Werte ist im Verlauf der ontogenetischen Einstellungsentwicklung in der späten Kindheit, zumindest nach einigen Autoren, zu erwarten (Nesdale, 2000; s. Kapitel II.1. Der Entwicklungsverlauf von Vorurteilen und Stereotypen). Somit könnte es mittels PARTS gelungen sein, diesen Wiederanstieg der negativen gruppenbezogenen sozialen Einstellungen in der Interventionsgruppe zu stoppen oder doch zumindest abzumildern oder zu verzögern.

Die Wirksamkeit auf den übrigen distalen Zielkriterien, also auf den interpersonalen sozialen Einstellungen sowie auf der sozialen Kompetenz, muss als begrenzt eingestuft werden. Es finden sich kleine Beschleunigungseffekte bei der Subskala Intoleranz und im prosozialen Verhalten, eingeschätzt durch die Klassenlehrer der Untersuchungsteilnehmer. Somit zeigt sich hier, zumindest kurzfristig, ein Effekt von PARTS auf die allgemeine soziale Kompetenz der Kinder und das auf einem Einstellungsmaß und einem Verhaltensmaß durch Fremdeinschätzung. Für eine dauerhafte Förderung dieser wichtigen Bereiche bedarf es vermutlich einer deutlich stärkeren Interventionsintensität (vgl. Johnson et al., 1994, 1995). Zudem sind diese sozialen Einstellungen und Verhaltensweisen der

Kinder auch einer normalen Weiterentwicklung und Reifeprozessen unterworfen, so dass es nicht verwundert, dass die Kontrollgruppe den kleinen Vorsprung der Interventionsgruppe wieder aufholen konnte.

Differentielle Evaluation:

Betrachtung spezifischer Einstellungskomponenten

Es wurden bei der Evaluation von PARTS verschiedene Einstellungskomponenten mit kognitivem, affektivem und konativem Schwerpunkt separat erfasst und ausgewertet. Somit können spezifische Aussagen über die Wirksamkeit des Programms auf einzelne Einstellungskomponenten getroffen werden und eine umfassende Wirksamkeit des Programms beurteilt werden (s.o. Distale Wirksamkeit).

Risikogruppen

In der Literatur ist es gängige Vorgehensweise als Zielgruppe einer Intervention die Kinder auszuwählen, welche ein hohes Ausmaß an negativen gruppenbezogenen sozialen Einstellungen aufweisen (Katz, 1973; Katz & Zalk, 1978; Aboud & Fenwick, 1999). Die explorative Untersuchung derartiger Risikogruppen ergab bei der Evaluation von PARTS ein ähnliches Ergebnismuster wie in der Gesamtstichprobe. Lediglich in der Effektstärkengröße und der statistischen Bedeutsamkeit der Interventionseffekte unterscheiden sich die Befunde von der Gesamtauswertung. Tendentiell scheint PARTS bei Risikogruppen ähnlich zu wirken wie bei der Normgruppe (s. Anhang Risikogruppen 25).

Generalisierung

Leider ist es nicht gelungen, diese positiven Effekte auf eine andere Fremdgruppe zu generalisieren. Abgesehen von kleinen Effekten bei affektiven Einstellungsmaßen zeigen sich hier kaum Interventionseinflüsse. Dies könnte einerseits eine Frage der notwendigen Intensität des Trainingsprogramms sein. Andererseits könnte die Ursache für die nicht vorhandene Generalisierung der Interventionseffekte in einem „Bodeneffekt“ der Einstellungsverzerrungen gegenüber der chinesischen Fremdgruppe liegen (v.a. auf der Verkuyten-Skala und MRA-Eigengruppenbevorzugung). Hier finden sich bei der Interventionsgruppe schon zum Prätest Werte auf einem nahe Null liegenden Niveau – d.h. es liegt so gut wie kein

Bias gegenüber Chinesen vor. Ein weiteres Absinken der Werte würde eine Bevorzugung der Fremdgruppe gegenüber der Eigengruppe bedeuten. Dies würde jedoch einer natürlichen und auch selbstwertfördernden Tendenz der Eigengruppenbevorzugung zuwiderlaufen. Das gehört nicht zu den Zielen von PARTS.

Im Bereich des prosozialen Verhaltens ist der Interventionseffekt kontextspezifisch geblieben. Leider scheint es hier nicht gelungen zu sein, den Interventionseffekt auf dem prosozialen Verhalten der Kinder im Schulkontext, eingeschätzt durch die Klassenlehrer der Kinder, auf einen anderen Kontext, zu Hause, zu generalisieren. Dies könnte an der mangelnden Intensität der Intervention liegen. Es könnte aber auch ein Effekt der mangelnden kindzentriertheit eines manualisierten Interventionsprogramms sein, wie es bei PARTS der Fall ist. Durch das zu geringe Eingehen auf die einzelnen Bedürfnisse der Kinder, könnte eine Verankerung der Interventionsinhalte in der restlichen Lebensumwelt der Teilnehmer nicht geglückt sein (s. Kapitel II.4. Schlussfolgerungen für PARTS). Schließlich könnte dieses Ergebnis auch ein weiterer Hinweis für die kontextspezifische Wirkung solcher Präventionsprogramme sein.

Allgemeine Wirksamkeit

Die vorgefundenen Interventionseffekte wurden mittels Kovarianzanalysen auf ihre Stabilität unter Einbezug zentraler Kovariaten geprüft. Dabei zeigten sich keine signifikanten Effekte der Kovariaten Geschlecht, Alter, Intelligenz und Entwicklungskontext auf die beschriebenen Evaluationsergebnisse. Genauso sprechen die zwar statistisch bedeutsamen aber nur kleinen Zusammenhänge von oben genannten Personenvariablen mit den Implementationsparametern für eine allgemeine Wirksamkeit von PARTS. Anders ausgedrückt ist die Wirksamkeit von PARTS unabhängig von, den hier geprüften, differentiellen Einflussfaktoren.

Zeitliche Perspektive

Hier ergibt sich ein gemischtes Bild. Interpersonale soziale Einstellungen und allgemeines Sozialverhalten konnten vor allem kurzfristig gefördert werden³³. Es

³³ Dabei muss an dieser Stelle erwähnt werden, dass der Posttest der PARTS-Untersuchung erst zwei Monate nach Beendigung des Trainingsprogramms durchgeführt wurde. Im Vergleich dazu wurde in

handelt sich hier um einen Beschleunigungseffekt. Die gruppenbezogenen sozialen Einstellungen hingegen konnten größtenteils nur verzögert durch die Intervention beeinflusst werden. Hier liegt also ein Sleeper-Effekt vor. Dieser wird durch zwei Prozesse qualifiziert: Zum einen findet zwischen zweitem und drittem Messzeitpunkt ein leichter Abfall der Biaswerte in der Interventionsgruppe statt, zum anderen zeigt sich in der Kontrollgruppe ein leichter Wiederanstieg der Einstellungsverzerrungen in diesem Vergleichszeitraum (s.o.). Dieses Muster findet sich auf verschiedenen Messinstrumenten, wie den Skalen des MRA oder auch des IBM.

vielen Studien aus der Literatur der Posttest direkt nach der Intervention angesetzt. Dies sollte bei der Beurteilung der kleinen „kurzfristigen“ Interventionseffekte bei PARTS beachtet werden.

V.7. Grenzen und Möglichkeiten der Arbeit

In jeder Forschungsarbeit ergeben sich aus unterschiedlichen Gründen Begrenzungen oder Mängel, die bei der Bewertung der Arbeit und der Interpretation der Ergebnisse berücksichtigt werden müssen. So auch in der vorliegenden Arbeit. Daher werden im Folgenden die zentralen Grenzen, aber auch die besonderen Qualitäten dieser Arbeit diskutiert.

Ein erstes Problem der vorliegenden Studie liegt in der Datenlage. Es wurden einige Skalen zur Erfassung der gruppenbezogenen sozialen Einstellungen gegenüber der russischen Zielfremdgruppe zu den ersten beiden Messzeitpunkten nicht erfasst. Dies hat verschiedene Gründe. Zum einen beruht diese Entscheidung auf den Zielstellungen von PARTS, positive gruppenbezogene soziale Einstellungen allgemein sowie Toleranz zu fördern. Das Programm sollte eine positive Wirkung auf die basalen gruppenunspezifischen Einstellungen entfalten und dieser Effekt sollte sich auf gruppenbezogene soziale Einstellungen gegenüber verschiedensten Fremdgruppen auswirken. Dieses Ziel wurde in der Form nicht erreicht. Eine Generalisierung der Effekte auf die chinesische Fremdgruppe fand bisher nicht statt. Ein weiterer Grund für die fehlenden Einstellungsmessungen liegt in der Vermeidung von Testwiederholungseffekten, die mögliche Interventionseffekte überdecken könnten. Und schließlich lag der Umfang der Testbatterie in den ersten Messungen an der Grenze der Belastbarkeit und Konzentrationsfähigkeit der Kinder, weshalb auch hier Abstriche bei den Maßen gemacht werden mussten. Somit fehlt eine wichtige Datengrundlage zur Beurteilung der Interventionseffekte, da nicht mit hundertprozentiger Sicherheit gewährleistet werden kann, ob die auf den erst zu t3 eingesetzten Skalen gefundenen Interventionseffekte nicht auf Niveauunterschieden beruhen, die es möglicherweise schon vor der Intervention gab. Allerdings stützt die vorhandene Datenlage diese Alternativerklärung nicht. So finden sich abgesehen von der Verkuyten-Skala gegenüber der Gruppe chinesischer Kinder keine bedeutsamen Prätestunterschiede zwischen den beiden Untersuchungsgruppen. Weiterhin finden sich besagte Interventionseffekte sowohl auf allen nur zu t3 eingesetzten Skalen als auch auf den MRA-Skalen, die auch schon zu t1 und t2 zum Einsatz kamen.

Schließlich wurden die Untersuchungsgruppen zumindest klassenweise randomisiert zusammengestellt, um die Wahrscheinlichkeit möglicher systematischer

Prätestunterschiede zu minimieren. Somit ergibt sich, unter Berücksichtigung der Datenlage und der Zusammenstellung der Untersuchungsgruppen, ein sinnvolles Bild. Man kann mit hoher Wahrscheinlichkeit von validen Interventionseffekten auf den gruppenbezogenen sozialen Einstellungen gegenüber der russischen Zielfremdgruppe ausgehen.

Ein weiterer Mangel in der vorliegenden Datenlage ist das Fehlen eines proximalen Effektmaßes für die Stellvertretender-Kontakt-Intervention. Hier hätte ein Maß zu einem bereits identifizierten Mediatorprozess wie Inclusion of others in the self (vgl. Wright et al., 1997) oder einen Manipulationscheck über die wahrgenommene Fremdgruppenhomogenität (vgl. Cameron et al., 2006a) eingesetzt werden können. Da die Studien zum Stellvertretenden-Kontakt der Arbeitsgruppe um Cameron sowie Kontaktstudien allgemein (Pettigrew & Tropp, 2006; Tropp & Prenovost, 2008) aber meist sehr erfolgreich evaluiert wurden, konnte davon ausgegangen werden, dass bei gut kontrollierter Umsetzung dieser Interventionsbestandteil auch bei PARTS gut funktioniert. Zudem lässt sich aus der Befundlage der gruppenbezogenen sozialen Einstellungen schließen, dass die Stellvertretender-Kontakt-Intervention erfolgreich implementiert werden konnte, da sich hauptsächlich auf der hier verwendeten Zielfremdgruppe Interventionseffekte zeigten.

Weiterhin fehlen Daten der Lehrereinschätzung des Sozialverhaltens der Teilnehmer zum dritten Messzeitpunkt. Dies geschah, um die sehr positive Zusammenarbeit mit den beteiligten Lehrern und Schulleitern nicht durch eine zu große zusätzliche Belastung durch die Studie zu gefährden. Zudem wurden die Lehrer im Verlauf des Projektes zum vierten Messzeitpunkt wieder befragt, so dass die hier fehlenden Daten zur Gesamteinschätzung der Interventionseffekte nicht so schwer ins Gewicht fallen. Die Ergebnisse hierzu werden jedoch an anderer Stelle berichtet werden.

Ein anderes Problem der vorliegenden Arbeit liegt in der mangelnden Qualität einiger proximaler Effektmaße. Zum einen sind die Maße, abgesehen vom Wissenstest nicht genügend auf die Interventionsinhalte abgestimmt. Somit können sie ihren eigentlichen Zweck der direkten Erfolgsmessung der eingesetzten Interventionsbestandteile nur bedingt erfüllen. Zum anderen weisen die proximalen Maße zu den sozial-kognitiven Kompetenzen, abgesehen vom Klassifikationstest, eine schlechte interne Konsistenz auf, was eine reliable und valide Messung in

diesem Bereich sehr erschwert, ganz zu schweigen vom Auffinden möglicher, vermutlich kleiner Interventionseffekte. Hierzu muss angemerkt werden, dass die vorliegende Arbeit im deutschen Sprachraum eine Art Pioniercharakter hat, da hier bisher keine Studie zu diesem Thema in einer derartigen Breite durchgeführt wurde. So ist es, trotz umfassender Recherchen sowie Testungen der Instrumente in einer Pilotstudie, leider nicht gelungen, adhoc zuverlässigere Instrumente zu diesem Bereich zu gestalten.

Auch bei der Auswahl der distalen Erfolgsmaße muss bei der vorliegenden Studie eine Einschränkung zugestanden werden. Es wurden bei der PARTS-Studie ausschließlich explizite Vorurteilsmaße verwendet. Gruppenbezogene soziale Einstellungen können von impliziten (unbewussten) oder expliziten (bewussten) Prozessen bestimmt werden (Greenwald & Banaji, 1995). Implizite und explizite Einstellungen können dabei sowohl kongruent als auch inkongruent innerhalb einer Person sein. Devine (1989) schlägt hierzu vor, dass implizite Einstellungen gut etablierte Überzeugungen beinhalten, die in der Kindheit erworben werden und explizite Einstellungen später und bewusster konstruiert werden. Sie hängen nur bis zu einem gewissen Grad positiv miteinander zusammen (Dovidio et al, 1996, 2004). Es war in der PARTS-Studie aus organisatorischen Gründen nicht möglich, auch implizite Einstellungsmaße zu erheben³⁴. In diesem Zusammenhang wird häufig die Kritik geäußert, dass durch Präventionsprogramme lediglich explizite gruppenbezogene soziale Einstellungen beeinflusst werden, indem die Kinder lernen, sich sozial erwünscht zu verhalten. Währenddessen blieben die impliziten gruppenbezogenen sozialen Einstellungen aber bestehen. Hierzu sei angemerkt, dass auch das Erlernen einer sozial erwünschten Norm zu gruppenbezogenen sozialen Einstellungen und zum sozialen Umgang miteinander vom Autor als ein Erfolg der Intervention angesehen wird.

Ein weiteres Manko der vorliegenden Arbeit besteht in der fehlenden Identifikation der Wirksamkeitsprozesse des Trainingsprogramms. Da es sich bei PARTS um ein multimodales Interventionsprogramm handelt, könnten sowohl einzelne Interventionsbestandteile, als auch alle gemeinsam, für die Wirksamkeit des Programms verantwortlich sein. Welche Prozesse allerdings genau ablaufen, kann mit dem vorliegenden Untersuchungsdesign mit nur zwei Untersuchungsgruppen nicht getestet werden. Dazu hätten verschiedene Interventionsgruppen mit den

³⁴ Hierzu wären Verfahren mit Reaktionszeitparadigmen notwendig gewesen, die sich bei einer aufsuchenden Studie mit Testungen in Schulen kaum realisieren lassen.

einzelnen Programmbestandteilen trainiert werden müssen, die dann bei der Auswertung der Daten miteinander verglichen worden wären. Dies hätte jedoch den Nachteil gehabt, die Stichprobe auf verschiedene Gruppen aufteilen zu müssen, und somit kleinere Interventionseffekte auf Grund der dann fehlenden Testpower nicht identifizieren zu können. Auch statistische Modelle zur Identifikation der Wirkprozesse lassen sich bei der vorliegenden Datenlage nicht gut anwenden. Zum einen gibt es in der Forschung noch keine konkreten Hinweise, wie derartige Mediationsmodelle genau aussehen könnten. Somit wäre eine derartige Analyse bestenfalls explorativ. Zum anderen erschweren die teilweise fehlenden (Stellvertretender-Kontakt), teilweise problematischen (sozial-kognitive Kompetenzen) proximalen Effektmaße das Auffinden möglicher Zusammenhänge. Schließlich würden ausführliche Analysen zu den Wirksamkeitsprozessen den Umfang dieser Arbeit sprengen. Daher beschränkt sich die Arbeit auf Aussagen über die Wirksamkeit des Interventionsprogramms.

Ein letzter Kritikpunkt der vorliegenden Arbeit liegt in einer Gefährdung der Objektivität der Untersuchung. Diese liegt darin begründet, dass sowohl die Programmentwicklung, als auch die Programmdurchführung und die Evaluation des Programms von ein und derselben Arbeitsgruppe durchgeführt wurden. Es ist bekannt, dass die Effektivität von derartigen Pilotstudien zur Wirksamkeit von Interventionsprogrammen immer leicht überschätzt werden, wenn die Programme von den Programmentwicklern selbst durchgeführt und evaluiert werden (Beelmann, 2006). Auf Grund der relativ dünnen Personaldecke des Forschungsprojekts ließ sich eine Überlappung der Mitarbeiter in den verschiedenen Aufgabenbereichen Entwicklung, Durchführung und Evaluation leider nicht vermeiden. Allerdings wurde bei den Erhebungen darauf geachtet, dass die PARTS-Trainer nicht ihre eigenen Interventionsteilnehmer befragten, um so möglichen Effekten der sozialen Erwünschtheit aus dem Wege zu gehen.

Die vorliegende Arbeit weist allerdings auch einige Vorzüge auf, die es lohnen, hervorgehoben zu werden. Zunächst verfügt die Untersuchung, auf der diese Arbeit basiert, über ein sehr gutes und aufwendiges Design. Das Prä-Post-Kontrollgruppendesign, gepaart mit der klassenweise randomisierten Zuweisung der Teilnehmer zu den Untersuchungsgruppen, gewähren dieser Studie eine sehr hohe interne Validität. Somit können Alternativerklärungen für die gefundenen

Interventionseffekte, abgesehen von den Effekten auf nicht zu allen Messzeitpunkten verwendeten Skalen, weitestgehend ausgeschlossen werden. Weiterhin können durch die Follow-up-Erhebung auch längerfristige oder verzögerte Interventionseffekte identifiziert bzw. abgesichert werden. International gibt es nur sehr wenige Studien mit einem derart aufwendigen und langfristigen Untersuchungsdesign inklusive Follow-up. National ist die vorliegende Arbeit dahingehend einzigartig.

Des Weiteren verfügt die PARTS-Studie über eine sehr große repräsentative Stichprobe. Dies führt zu einer hohen externen Validität und Generalisierbarkeit der Aussagen über die Umsetzbarkeit und Wirksamkeit von PARTS. Zudem wird durch diese relativ große Stichprobe auch das Auffinden und statistische Absichern kleinerer Interventionseffekte möglich. Weiterhin konnte durch eine aufwendige Stichprobenpflege erreicht werden, dass es nur eine minimale Ausfallrate der Teilnehmer über die drei Messzeitpunkte gab. Das bedeutet, dass es nur sehr wenig fehlende Daten zu den einzelnen Messzeitpunkten gibt. Somit sind zuverlässige Aussagen zur Entwicklung der Teilnehmer über alle drei Messzeitpunkte möglich. All dies spricht für eine sehr hohe Datenqualität der PARTS-Untersuchung.

Bei der PARTS-Studie wurde eine umfangreiche Implementationskontrolle durchgeführt. Somit können detaillierte Aussagen über die Umsetzbarkeit jeder Interventionseinheit, der einzelnen Interventionsmethoden sowie der Passung des Programms für bestimmte Adressaten getroffen werden. Zudem wird die Qualität der Programmdurchführung dokumentiert und sichergestellt. Eine solche Implementationskontrolle findet sich bei bisherigen Studien zur Vorurteilsprävention sehr selten.

Weiterhin verfügt die vorliegende Untersuchung über ein sehr breites Spektrum an Effektmaßen. Es wurden sowohl proximale als auch distale Effektmaße zur Absicherung der Wirksamkeit des Programms eingesetzt. Bei den distalen Effektmaßen wurden unterschiedliche Aspekte von gruppenbezogenen sozialen Einstellungen gegenüber verschiedenen Fremdgruppen erfasst, ebenso wie allgemeine soziale Einstellungen und Einschätzungen des Sozialverhaltens der Kinder. Dazu wurden verschiedene Informanten (Kinder, Lehrer, Eltern) befragt. Somit bilden die vorliegenden Befunde ein sehr breites Spektrum von Kompetenzen, sowie verschiedene Dimensionen von Einstellungen und Verhaltensweisen ab. Dies

ermöglicht sehr differenzierte und sehr zuverlässige Aussagen über die Wirksamkeit von PARTS.

Insgesamt kann festgehalten werden, dass es weltweit so gut wie keine Evaluationsstudien zum Thema Vorurteilsprävention bei Kindern mit einer derart hohen Qualität gibt, wie dies bei der PARTS-Studie der Fall ist.

V.8. Implikationen für Forschung und Praxis

Aus den Ergebnissen der vorliegenden Arbeit lassen sich verschiedene Schlüsse für die Forschung und für die Anwendung von Trainingsprogrammen zur Prävention negativer gruppenbezogener sozialer Einstellungen ziehen. So wartet zunächst die offen gebliebene Frage der Wirkprozesse hinter den gefundenen Interventionseffekten auf ihre Beantwortung. Diese Frage ist allerdings von zentraler Bedeutung, wenn es darum geht, welche Interventionsform verwendet werden muss, wenn ein umfassendes ganzheitliches Ergebnis bei der Beeinflussung von gruppenbezogenen sozialen Einstellungen erreicht werden soll. Dazu kann an dieser Stelle festgehalten werden, dass das multimodale Präventionsprogramm PARTS einen, zwar verzögerten und kleinen, dafür aber umfassenden Effekt auf allen erfassten Einstellungsfacetten gegenüber der Zielfremdgruppe gezeigt hat. Ob es dafür genügt hätte, eine reine Kontakt-, oder Stellvertretender-Kontakt-Intervention durchzuführen, wie die hier auftretenden facettenübergreifenden aber rein zielfremdgruppenbezogenen Effekte nahe legen könnten, oder ob man multimodal vorgehen muss, um alle Facetten von gruppenbezogenen sozialen Einstellungen wirksam ansprechen zu können, bleibt offen. Dovidio und Kollegen (2004) plädieren in ihrem Artikel über die Wirkprozesse hinter Einstellungsmodifikationen für ein multimodales Vorgehen, um verschiedene Einstellungsbereiche mit verschiedenen Methoden optimal ansprechen zu können (s. Wirkprozesse im Kapitel II.3. Empirische Studien). Ein weiteres Argument für das multimodale Vorgehen ist in jedem Fall der gefundene Beschleunigungseffekt in der Interventionsgruppe zu interpersonalen sozialen Einstellungen (Intoleranz) und positivem Sozialverhalten (prosoziales Verhalten). Diesen Effekt hätte es bei einem reinen Kontaktprogramm vermutlich kaum gegeben. Fest steht allerdings auch, dass Kontakt-Interventionen ein elementarer Bestandteil zur Einstellungsmodifikation sind und bleiben (Pettigrew & Tropp, 2006; Tropp & Prenovost, 2008). Dafür sprechen auch die Ergebnisse der vorliegenden Arbeit. Die Frage der Wirkprozesse ist in jedem Fall eine wichtige und lohnende Forschungsfrage, die dringend weiter verfolgt werden muss, um sowohl die Forschung als auch die Praxis der Vorurteilsprävention weiter voranbringen zu können.

Eine andere Frage ist die nach der notwendigen Intensität und Dauer einer wirksamen Intervention. PARTS konnte zeigen, dass mit einem Programm mit 15 Einheiten über 7 Wochen Netto-Interventionsdauer durchaus kleine Effekte erzielt werden können. Genauso deutlich wurde aber, dass die langfristige Förderung verschiedener sozial-kognitiver Kompetenzen mehr Interventionsintensität benötigt. Bisher gibt es nur sehr grobe Richtlinien und Erfahrungswerte zu dieser Frage. Um eine optimale, effiziente Intervention planen zu können, sind genaue Informationen zur benötigten Dauer und Intensität einzelner Interventionsformen unumgänglich. Daher wäre es wünschenswert, wenn auch in diesem Bereich eine vertiefte Forschung ansetzen würde.

Eine weitere zentrale Implikation dieser Arbeit ergibt sich aus der Breite der vorgefundenen Effekte, gepaart mit deren relativ geringer Größe. Hier scheint es so zu sein, dass durch die Aufteilung der Interventionsressourcen auf viele verschiedene Bereiche zwar viele verschiedene Einstellungsfacetten angesprochen werden, dafür aber die notwendige Intensität in den einzelnen Bereichen zur stärkeren Förderung eingebüßt wird. Ein multimodales Trainingsprogramm mit einer höheren Intensität der einzelnen Interventionsbereiche wäre für eine dauerhafte und praxisrelevante Förderung positiver Intergruppenbeziehungen wünschenswert.

Eine weitere Schlussfolgerung aus der vorliegenden Arbeit ist die Möglichkeit, auch in homogenen Umwelten erfolgreiche Vorurteilsprävention zu betreiben. In den 15 von PARTS besuchten Grundschulklassen beschränkte sich die Anzahl der Kinder mit Migrationshintergrund auf 9. Somit sind hier viele kontaktbasierte Interventionsformen nicht möglich. PARTS lässt sich in solchen homogenen Umwelten jedoch problemlos durchführen. Dabei sollte zuvor geprüft werden, welche sozialen Gruppen und gruppenbezogenen sozialen Einstellungen in dem entsprechenden sozialen Kontext vorherrschen. Dies sollte unbedingt berücksichtigt werden, um eine relevante Zielfremdgruppe für die jeweiligen Adressaten des Programms auswählen zu können.

Dies führt gleich zur nächsten Schlussfolgerung aus den Ergebnissen der vorliegenden Arbeit. So ist es nicht gelungen, die Interventionseffekte gegenüber der Zielfremdgruppe auf eine zweite Fremdgruppe zu übertragen. Dies könnte daran liegen, dass die Interventionsintensität zu gering war. Oder dieses Ergebnis beruht auf einer Art Bodeneffekt, da die sozialen Einstellungen gegenüber der Gruppe der Chinesen bereits zum ersten Messzeitpunkt in beiden Untersuchungsgruppen sehr

positiv waren. Oder Einstellungsveränderungen finden generell nur gruppenspezifisch statt. In jedem Fall kann festgehalten werden, dass es sinnvoll ist, sich bei der Planung einer Intervention zur Prävention negativer gruppenbezogener sozialer Einstellungen klar zu werden, gegenüber welcher Gruppe diese Einstellungen modifiziert werden sollen. Weiterhin gibt es den Hinweis auf eine kontextspezifische Wirkung bei der Förderung des prosozialen Verhaltens. Hier fanden sich im schulischen Kontext zumindest kurzfristig positive Interventionseffekte, nicht jedoch im häuslichen Kontext. Dies könnte damit zusammenhängen, dass soziale Normen zum Sozialverhalten in einem sozialen Kontext erlernt und zunächst auch nur dort angewendet werden. Diese Normen scheinen hier nicht auf den anderen Kontext übertragen worden zu sein, oder die Eltern der Kinder waren in ihren Einschätzungen nicht sensitiv genug, um auch kleine Veränderungen festzustellen. Hier scheint jedenfalls ein zielgruppenspezifisches und kontextspezifisches Vorgehen erfolgversprechender, als eine unspezifische allgemeine Zielsetzung der positiven Beeinflussung gruppenbezogener sozialer Einstellungen und positiven Sozialverhaltens, wie sie bei PARTS verfolgt wird.

Ein letzter wichtiger Punkt betrifft die Auswahl des Zeitfensters für Vorurteilspräventionsprogramme. Hier hat sich bei PARTS die Übergangszeit von der mittleren zur späten Kindheit als sehr günstig erwiesen. Denn während sich in der Kontrollgruppe ein Wiederansteigen der gruppenbezogenen sozialen Einstellungsverzerrungen zeigte, konnte bei der Interventionsgruppe ein weiterer Abfall der Werte bzw. eine Stabilität verzeichnet werden. Somit scheint hier der normative Wiederanstieg der Einstellungsverzerrungen (s. Kapitel II.2.1. Entwicklungsverlauf von Vorurteilen und Stereotypen) verhindert, oder zumindest verzögert bzw. verringert worden zu sein. Die wissenschaftlich fundierte Auswahl des für Vorurteilspräventionen sensiblen Zeitfensters, hat somit hervorragend funktioniert. Dabei wurde jedoch zusätzlich auf eine altersentsprechende inhaltliche Schwerpunktsetzung mit einem sozial-kognitiven Programmfokus geachtet (vgl. Teichman & Bar-Tal, 2008). Dies kann als Hinweis für die Durchführung anderer sozial-kognitiver Vorurteilspräventionsprogramme in der mittleren bis späten Kindheit dienen, oder als Forderung für eine entwicklungspsychologische Fundierung von Interventionen allgemein verstanden werden.

VI. Zusammenfassung

Ziel dieser Arbeit ist die Evaluation eines multimodalen Präventionsprogramms zur Förderung von Intergruppenbeziehungen bei Grundschulkindern. Kernidee hierzu ist die Prävention negativer gruppenbezogener sozialer Einstellungen, sowie die Förderung von Toleranz und sozialer Kompetenz. Die für die Entwicklung eines solchen Programms notwendigen theoretischen Grundlagen, wie Begriffsdefinition, Entwicklungsverlauf und –theorien zu Vorurteilen, sowie die zentralen Evaluationsstudien zu diesem Thema, werden in der Arbeit dargelegt. Aus dieser theoretischen Basis lassen sich verschiedene Forderungen an ein wissenschaftlich fundiertes Programm, wie ein multimodales, präventives und entwicklungsangemessenes Vorgehen, ableiten.

Das daraus hervorgehende Programm PARTS, Programm zur Förderung von Akzeptanz, Respekt, Toleranz und Sozialer Kompetenz, beruht auf drei, bereits als effektiv evaluierten, Interventionsmethoden zur Prävention von Vorurteilen und Diskriminierung: Stellvertretender-Kontakt, wissensbasierte Intervention und Förderung sozial-kognitiver Kompetenzen. Dabei verfügt PARTS über einen altersentsprechenden sozial-kognitiven Fokus.

Das Programm wird mit einem aufwendigen Prä-Post-Kontrollgruppendesign mit Follow-up an einer ausreichend großen Stichprobe Thüringer Grundschüler durchgeführt und evaluiert. Dazu kommen verschiedene proximale und distale Effektmaße zum Einsatz, um die Wirksamkeit von PARTS umfassend abzubilden. Des Weiteren wird PARTS bei der Durchführung einer genauen Implementationskontrolle unterzogen.

Die Ergebnisse zeigen, dass sich PARTS sehr gut in den teilnehmenden Schulklassen implementieren lässt. Die Beteiligung der Kinder am Programm ist sehr gut, das Störverhalten sehr gering und die Freude an den Aufgaben und Spielen sehr groß. Zudem gibt es nur eine kleine Gruppe von Kindern, die dauerhaft Störverhalten zeigen.

Die Ergebnisse der proximalen Effektmaße deuten darauf hin, dass vor allem die wissensbasierte Interventionsmethode bei PARTS erfolgreich ist, während die Förderung der sozial-kognitiven Kompetenzen nur bedingt Wirkung zeigt.

Die Ergebnisse der distalen Effektmaße weisen zudem auf die Wirksamkeit der Stellvertretender-Kontakt-Intervention hin. Es zeigen sich auf verschiedenen Maßen zu gruppenbezogenen sozialen Einstellungen gegenüber der russischen Zielfremdgruppe Entwicklungsunterschiede zwischen den Untersuchungsgruppen. Diese Zielfremdgruppe kommt ausschließlich bei der Stellvertretender-Kontakt-Intervention zum Einsatz, was ein Hinweis für deren Wirksamkeit ist.

Die Interventionseffekte gegenüber der russischen Zielfremdgruppe treten hauptsächlich verzögert, zwischen dem zweiten und dem dritten Messzeitpunkt auf. Sie lassen sich auf Einstellungsmaßen mit kognitiven, affektiven und konativen Schwerpunkten finden. Dies spricht für eine umfassende Wirksamkeit des multimodalen Programms auf verschiedenen Facetten gruppenbezogener sozialer Einstellungen. Die gefundenen Interventionseffekte sind klein, aber praktisch und statistisch bedeutsam. Die Effekte finden sich nicht bei der chinesischen Fremdgruppe.

Weiterhin findet sich ein Beschleunigungseffekt in der Abnahme der Intoleranz der Interventionsteilnehmer, sowie eine zumindest kurzfristige Zunahme im prosozialen Verhalten der PARTS-Teilnehmer im Vergleich zur Kontrollgruppe. Insgesamt ist PARTS, zwar zielgruppenspezifisch aber umfassend, wirksam in der Prävention von negativen gruppenbezogenen sozialen Einstellungen sowie in der Förderung interpersonaler Einstellungen und prosozialen Verhaltens.

Diese Ergebnisse werden vor dem aktuellen Forschungshintergrund diskutiert und die Grenzen und Möglichkeiten der vorliegenden Arbeit erläutert. Abschließend werden Implikationen für die weitere Forschung, sowie für die praktische Arbeit zur Förderung von Intergruppenbeziehungen bei Kindern entwickelt.

VII. Literatur

- Aarim-Herlot, N. (2003). *Chinese immigrants, African Americans, and racial anxiety in the United States 1848- 82*. Urbana, IL: University of Illinois Press.
- Aboud, F. E. (1988). *Children and prejudice*. New York: Blackwell.
- Aboud, F. E. (2002, June). *Changing racial attitudes of young children with antibias messages*. Paper presented at the Symposium talk at the Society for the Psychological Study of Social Issues
- Aboud, F. E. (2003). The formation of in-group favoritism and out-group prejudice in young children: Are they distinct attitudes? *Developmental Psychology*, 39, 48- 60.
- Aboud, F. E. (2008). A social-cognitive developmental theory of prejudice. In S. M. Quintana & C. Mckown (Eds.), *Handbook of race, racism and the developing child*. Hoboken, NJ: John Wiley & Sons.
- Aboud, F. E., & Amato, M. (2001). Developmental and socialization influences on intergroup bias. In R. Brown & S. Gaertner (Eds.), *Blackwell handbook in social psychology*. New York: Blackwell.
- Aboud, F. E., & Doyle, A. B. (1996a). Parental and peer influences on children's racial attitudes. *International Journal of Intercultural Relations*, 20, 371- 383.
- Aboud, F. E., & Doyle, A. B. (1996b). Does talk of race foster prejudice or tolerance in children? *Canadian Journal of Behavioural Science*, 28, 161-170.
- Aboud, F. E., & Fenwick, V. (1999). Exploring and evaluating school-based interventions to reduce prejudice. *Journal of Social Issues*, 55(4), 767-786.
- Aboud, F. E., & Levy, S. R. (2000). Interventions to reduce prejudice and discrimination in children and adolescents. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (269-293). Mahwah: Lawrence Erlbaum Associates.
- Abrams, D., Rutland, A., & Cameron, L. (2003). The development of subjective group dynamics: Children's judgments of normative and deviant in-group and out-group individuals. *Child Development*, 74, 1-17.
- Adorno, T., Frenkel-Brunswick, E., Levinson, D., & Sanford, R. (1950). *The authoritarian personality*. New York: Norton.
- Agnes, Sister Mary S. C. (1947). Influence of reading on the racial attitudes of adolescent girls. *Catholic Educational Review*, 45, 415- 420.

- Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63, 596-612.
- Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close relationships as including other in the self. *Journal of Personality and Social Psychology*, 60(2), 241-253.
- Ashmore, R. D., Deaux, K., & McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: Articulation and significance of multidimensionality. *Psychological Bulletin*, 130, 80-114.
- Assaf, R. (2006). *Perceptions and attitudes of Jewish children towards Arabs and Jews*. Unpublished Master's Thesis, Tel-Aviv University, Tel-Aviv, Israel (in Hebrew).
- Augoustinos, M., & Rosewarne, D.L. (2001). Stereotype knowledge and prejudice in children. *British Journal of Developmental Psychology*, 19, 143-156.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, New Jersey: Prentice Hall.
- Banks, A., & McGee-Banks, C. A. (2001). *Handboof of research on multicultural education*. San Francisco, CA: Jossey- Bass Inc.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bar-Tal, D. (1996). Development of social categories and stereotypes in early childhood: The case of the "Arab" concept formation, stereotype and attitudes by Jewish children in Israel. *International Journal of Intercultural Relations*, 20(3/4), 341- 370.
- Bar-Tal, D., & Teichman, Y. (2005). *Stereotypes and prejudice in conflict: Representations of Arabs in Israeli Jewish society*. New York: Cambrigde University Press.
- Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Erlbaum.
- Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imagining how you would feel. *Personality and Social Psychology Bulletin*, 23, 751-758.

- Beelmann, A. (2000). *Prävention dissozialer Entwicklungen: Psychologische Grundlagen und Evaluation früher kind- und familienbezogener Interventionsmaßnahmen*. Habilitationsschrift, Universität, Erlangen-Nürnberg.
- Beelmann, A. (2004). Förderung sozialer Kompetenzen im Kindergarten: Evaluation eines sozialen Problemlösetrainings zur universellen Prävention dissozialer Verhaltensprobleme. *Kindheit und Entwicklung*, 13(2), 113-121.
- Beelmann, A. (2006). Wirksamkeit von Präventionsmaßnahmen bei Kindern und Jugendlichen. Ergebnisse und Implikationen der integrativen Erfolgsforschung. *Zeitschrift für Klinische Psychologie und Psychotherapie* 35, 151-162.
- Beelmann, A. (2007). Elternberatung und Elterntraining. In F. Linderkamp & M. Grünke (Eds.), *Lern- und Verhaltensstörungen. Genese, Diagnostik, Intervention* (298- 311). Weinheim: Beltz Psychologie Verlags Union.
- Beelmann, A., & Heinemann, K. S. (in Vorbereitung). The prevention of prejudice in children and adolescents: A meta-analysis of psychological and educational training programs.
- Beelmann, A., Heinemann, K. S., & Saur, M. (2009). Interventionen zur Prävention von Vorurteilen und Diskriminierung. In A. Beelmann, & Jonas, K. J. (Ed.), *Diskriminierung und Toleranz: Psychologische Grundlagen und Anwendungsperspektiven* (435-461). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Beelmann, A., Jaursch, S., Lösel, F., & Stemmler, M. (2006). Frühe universelle Prävention von dissozialen Erziehungsproblemen: Implementation und Wirksamkeit eines verhaltensorientierten Elterntrainings. *Praxis der Rechtspsychologie*, 16(1/2), 120- 143.
- Beelmann, A., & Raabe, T. (2007). *Dissoziales Verhalten bei Kindern und Jugendlichen. Erscheinungsformen, Entwicklung, Prävention und Intervention*. Göttingen: Hogrefe.
- Beelmann, A., Saur, M., & Schulze, D. (2006). Präventionserfordernisse bei ideologiebasierter Delinquenz. In R. Egg (Ed.), *Extremistische Kriminalität: Kriminologie und Prävention* (165-194). Wiesbaden: Kriminologische Zentralstelle.

- Beelmann, A., Saur, M., Schulze, D., & Ziegler, P. (2007). PARTS - Programm zur Förderung von Akzeptanz, Respekt, Toleranz und Sozialer Kompetenz. Vorläufiges Trainingsmanual Universität Jena: Institut für Psychologie und DFG Forschergruppe Diskriminierung und Toleranz in Intergruppenbeziehungen.
- Beelmann, A., Saur, M., & Ziegler, P. (2010). Thüringer Studie zur Vorurteilsprävention und Toleranzentwicklung. Unveröffentlichter Projektbericht für das Thüringer Kultusministerium. Friedrich-Schiller-Universität Jena.
- Bertelsmann-Stiftung (Ed.). (2004). *Eine Welt der Vielfalt. Ein Trainingsprogramm des A WORLD OF DIFFERENCE-Institute der Anti-Defamation-League, New York, in der Adaptation für den Schulunterricht, Praxishandbuch für Lehrerinnen und Lehrer*. Gütersloh: Bertelsmann Stiftung.
- Bigler, R. S. (1995). The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development*, 66, 1072-1087.
- Bigler, R. S., Averhart, C. J., & Liben, L. S. (2003). Race and the workforce: occupational status, aspirations, and stereotyping among African American children. *Developmental Psychology*, 39(3), 572- 580.
- Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorisation and the formation of intergroup attitudes in children. *Child Development*, 68, 530- 543.
- Bigler, R. S., & Liben, L. S. (1992). Cognitive mechanisms in children's gender stereotyping: Theoretical and educational implications of a cognitive-based intervention. *Child Development*, 64, 1507-1518.
- Bigler, R. S., & Liben, L. S. (1993). A Cognitive-developmental approach to racial stereotyping and reconstructive memory in Euro-American children. *Child Development*, 64, 1507- 1518.
- Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, 39- 89). San Diego, CA: Elsevier Academic Press.
- Black-Gutman, D., & Hickson, F. (1996). The relationship between racial attitudes and social-cognitive development in children: An Australian study. *Developmental Psychology*, 32, 448-456.

- Blair, I. V. (2001). Implicit stereotypes and prejudice. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (359-374). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Blake, R., & Dennis, W. (1943). The development of stereotypes concerning the Negro. *The Journal of Abnormal and Social Psychology*, 38(4), 525- 531.
- Blascovich, J., Mendes, W. B., Hunter, S. B., Lickel, B., & Kowai-Bell, N. (2001). Perceiver threat in social interactions with stigmatized others. *Journal of Personality and Social Psychology*, 80(2), 253-267.
- Bobo, L. (1988). Group conflict, prejudice and the paradox of contemporary racial attitudes. In P. A. Katz & D. A. Taylor (Eds.), *Eliminating racism: Profiles in controversy* (85- 114). New York, NY, US: Plenum Press.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler*. Berlin/Heidelberg: Springer-Verlag.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge: University Press.
- Bowers, V., & Swanson, D. (1988). *More than meets the eye*. Vancouver, British Columbia: Pacific Educational Press.
- Branch, C. W., & Newcombe, N. (1986). Racial attitude development among young black children as a function of parental attitudes: A longitudinal and cross-sectional study. *Child Development*, 57, 712- 721.
- Brewer, M., & Miller, N. (1984). Beyond the contact hypothesis: Theoretical perspectives on desegregation. In N. Miller, & M. Brewer (Ed.), *Groups in conflict* (281-302). New York: Academic Press.
- Brewer, M. (1999). The psychology of prejudice: ingroup love or outgroup hate? *Journal of Social Issues*, 55, 429- 444.
- Brown, R. J. (1986). *Social Psychology* (2nd ed.). New York: Free Press.
- Bundesministerium des Innern (2009). *Verfassungsschutzbericht 2008*.
- Cameron, L., & Rutland, A. (2006). Extended contact through story reading in school: reducing children's prejudice toward the disabled. *Journal of Social Issues*, 62(3), 469-488.

- Cameron, L., Rutland, A., & Brown, R. (2007). Promoting children's positive intergroup attitudes towards stigmatized groups: Extended contact and multiple classification skills training *International Journal of Behavioral Development*, 31(5), 454-466.
- Cameron, L., Rutland, A., Brown, R., & Douch, R. (2006). Changing children's intergroup attitudes toward refugees: Testing different models of extended contact. *Child Development*, 77(5), 1208-1219.
- Carter, C., & Rice, C. L. (1997). Acquisition and manifestation of prejudice in children. *Journal of Multicultural Counseling and Development*, 25(3), 185-194.
- Chandler, M. J. (1973). Egocentrism and antisocial behavior: The assessment and training of social perspective-taking skills. *Developmental Psychology*, 9(3), 326-332.
- Chandler, M. J., Greenspan, S., & Barenboim, C. (1974). Assessment and training of role-taking and referential communication skills in institutionalized emotionally disturbed children. *Developmental Psychology*, 10(4), 546-553.
- Christiansen, S., & Janosch. (2006). *Gibt es hitzefrei in Afrika? So leben die Kinder dieser Welt*. München: Wilhelm Heyne Verlag.
- Clark, A., Hocevar, D., & Dembo, M. H. (1980). The role of cognitive development in children's explanations and preferences for skin color. *Developmental Psychology*, 16, 332- 339.
- Clark, K. B., & Clark, M. P. (1947). Racial identification and preference in Negro children. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology* (169-178). New York: Holt.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Coker, D. R. (1984). The relationships among gender concepts and cognitive maturity in preschool children. *Sex Roles*, 10, 19- 31.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Rand McNally, Chicago, Illinois.
- Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of black and white discrimination and prejudice: A literature review. *Psychological Bulletin*, 87(3), 546- 563.

- Deutsch, M., & Gerard, H. (1955). A study of normative and informational social influences on individual judgment. *Journal of Abnormal and Social Psychology*, 51(3), 629-636.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.
- Devine, P. G., & Monteith, M. J. (1993). The role of discrepancy-associated affect in prejudice reduction. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (317-344). San Diego, CA: Academic Press.
- Döpfner, M., Lorch, R., & Reihl, D. (1989). Soziale Informationsverarbeitung in Konfliktsituationen - eine empirische Studie an Vorschulkindern. *Zeitschrift für Pädagogische Psychologie*, 3, 239-248.
- Dovidio, J. F., Brigham, J. C., Johnson, B. T., & Gaertner, S. L. (1996). Stereotyping, prejudice and discrimination: Another look. In C. N. Macrae, C. Stangor & M. Hewstone (Eds.), *Stereotypes and stereotyping* (276-319). New York: Guilford.
- Dovidio, J. F., & Gaertner, S. L. (1998). On the nature of contemporary prejudice: The causes, consequences, and challenges of aversive racism. In J. L. Eberhardt & S. T. Fiske (Eds.), *Confronting racism: The problem and the response* (3-32). Thousand Oaks, CA, US: Sage Publications Inc.
- Dovidio, J. F., & Gaertner, S. L. (1999). Reducing prejudice: combating intergroup biases. *Current directions in psychological science*, 8(4), 101-105.
- Dovidio, J. F., Gaertner, S. L., Stewart, T. L., Esses, V. M., Vergert, M. T., & Hodson, G. (2004). From intervention to outcome: Processes in the reduction of bias. In W. G. Stephan & W. P. Vogt (Eds.), *Education programs for improving intergroup relations* (243- 265). New York: Teachers College Press.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2000). Reducing contemporary prejudice: Combating explicit and implicit bias at the individual and intergroup level. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (137-163). Mahwah, NJ, US: Lawrence Erlbaum Associate Publishers.
- Doyle, A. B., & Aboud, F. E. (1995). A longitudinal study of white children's racial prejudice as a social-cognitive development. *Merrill-Palmer Quarterly*, 41, 210-229.

- Duckitt, J. H. (1992a). Psychology and prejudice: A historical analysis and integrative framework. *American Psychologist*, 47, 1182- 1193.
- Duckitt, J. H. (1992b). *The social psychology of prejudice*. Westport, CT, US: Prager Publishers/Greenwood Publishing Group.
- Ehrlich, H. J. (1973). *The social psychology of prejudice*. New York: Wiley.
- Erikson, E. H. (1968). *Identity: Youth and crisis*. New York: Norton.
- Esses, V. M., & Dovidio, J. F. (2002). The role of emotions in determining willingness to engage in intergroup contact. *Personality and Social Psychology Bulletin*, 28, 1202-1214.
- Esses, V. M., Dovidio, J. F., Jackson, L. M., & Armstrong, T. L. (2001). The immigration dilemma: The role of perceived group competition, ethnic prejudice and national identity. *Journal of Social Issues*, 57(3), 389-412.
- Feddes, A., Noack, P., & Rutland, A. (2009). Direct and extended friendship effects on minority and majority children's interethnic attitudes: A longitudinal study. *Child Development*, 80, 377-390.
- Finlay, K. A., & Stephan, W. G. (2000). Improving intergroup relations: The effects of empathy on racial attitudes. *Journal of Applied Social Psychology*, 30(8), 1720-1737.
- Fishbein, H. D. (2002). *Peer prejudice and discrimination* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Fisher, F. (1965). *The influence of reading and discussion on the attitude of fifth graders toward American Indians*. Unpublished doctoral dissertation, University of California, Berkeley.
- Flavell, J., Botkin, P., Fry, C., Wright, J., & Jarvis, P. (1968). *The development of role-taking and communication skills in children*. New York: Wiley.
- Frederickson, G. M., & Knobel, D. T. (1980). History of prejudice and discrimination. In S. Thernstrom, A. Orlov & O. Handlin (Eds.), *Harvard encyclopedia of American ethnic groups* (829-847). Cambridge, MA: Belknap Press.
- Frindte, W., Neumann, J., Hieber, K., Knote, A., & Müller, C. (2001). Rechtsextremismus = "Ideologie plus Gewalt" - Wie ideologisiert sind rechtsextreme Gewalttäter? *Zeitschrift für Politische Psychologie*, 9(2-3), 81-98.

- Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, 95(1), 148-162.
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination and racism* (61- 89). San Diego, CA, US: Academic Press.
- Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. Hove, East Sussex: Psychology Press.
- Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology*, 57, 239-249.
- Gimmetstad, B. J., & DeChiara, E. (1982). Dramatic plays: A vehicle for prejudice reduction in the elementary school. *Journal of Educational Research*, 76, 45-49.
- Gonzalez, R., & Brown, R. (2003). Generalization of positive attitude as a function of subgroup and superordinate identifications in intergroup contact. *European Journal of Social Psychology*, 33, 195-214.
- Goodenow, C. (1993). Classroom belonging among early adolescent students: Relationships to motivation and achievement. *Journal of Early Adolescence*, 13, 21-43.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 38(5), 581-586.
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95-112.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4-27.
- Gregor, A. J., & McPherson, D. A. (1966). Racial attitudes among white and Negro children in a deep-south standard metropolitan area. *The Journal of Social Psychology*, 68(1), 95-106.
- Griffiths, J. A., & Nesdale, D. (2006). In-group and out-group attitudes of ethnic majority and minority children. *International Journal of Intercultural Relations*, 30, 735-749.

- Hamilton, D. L., & Troiler, T. K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (127-163). Orlando, FL: Academic Press.
- Heller, K. A., Kratzmeier, H., & Lengfelder, A. (1998). *Matrizen-Test-Manual* (Vol. 2). Göttingen: Beltz-Test.
- Hewstone, M., Brown, R. J. (1986). Contact is not enough: An intergroup perspective on the contact hypothesis. In M. Hewstone, & R. Brown (Ed.), *Contact and conflict in intergroup encounters* (1-44). Oxford, U.K.: Blackwell.
- Hewstone, M., & Brown, R. J. (2005). An integrative theory on intergroup contact. *Advances in experimental social psychology*, 37, 255-343.
- Hill, M. E., & Augustinos, M. (2001). Stereotype change and prejudice reduction: Short-and long-term evaluation of a cross-cultural awareness programme. *Journal of Community and Applied Social Psychology*, 11, 243-262.
- Holt, R. R., & Silverstein, B. (1989). On the psychology of enemy images: Introduction and overview. *Journal of Social Issues*, 45(2), 1-11.
- Horowitz, E. L. (1936). The development of attitude toward the Negro. *Archives of Psychology (Columbia University)*, 194, 47.
- Horowitz, E. L., & Horowitz, R. (1938). Development of social attitudes in children. *Sociometry*, 1, 301-338.
- Howell, M. (1973). *A study of the effects of reading upon the attitudes of fifth graders toward Mexican Americans*. Unpublished doctoral dissertation, Southern Illinois University, Carbondale.
- Ignatiev, N. (1995). *How the Irish became white*. New York: Routledge.
- Islam, M. R., Hewstone, M. (1993). Dimensions of contact as predictors of intergroup anxiety, perceived out-group variability, and out-group attitude: An integrative model. *Personality and Social Psychology Bulletin*, 19, 700-710.
- Jackson, E. P. (1944). Effects of reading upon attitudes toward the Negro race. *Library Quarterly*, 14, 47-54.
- Jelloun, T. B. (2000). *Papa, was ist ein Fremder?* Reinbek: Rowohlt.
- Johnson, D. W., & Johnson, R. T. (1981). Effects of cooperative and individualistic learning experiences on interethnic interaction. *Journal of Educational Psychology*, 73, 444-449.
- Johnson, D. W., & Johnson, R. T. (1991). *Teaching students to be peacemakers*. Edina, MN: Interaction Book Co.

- Johnson, D. W., & Johnson, R. T. (2000). The three cs of reducing prejudice and discrimination. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (239-268). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Johnson, D. W., & Johnson, R. T., Dudley, B. & Magnuson, D. (1995). Training elementary school students to manage conflict. *The Journal of Social Psychology*, 135(6), 673-686.
- Johnson, D. W., Johnson, R. T., Dudley, B., & Acikgoz, K. (1994). Effects of conflict resolution training on elementary school students. *The Journal of Social Psychology*, 134(6), 803-817.
- Johnston, L., & Hewstone, M. (1990). Intergroup contact: social identity and social cognition. In D. Abrams & M. A. Hogg (Eds.), *Social identity theory: Constructive and critical advances* (185-210). London: Harvester Wheatsheaf.
- Jugert, G., Kabak, S., & Notz, P. (2006). *Fit for Differences: Training interkultureller und sozialer Kompetenzen für Jugendliche*. Weinheim: Juventa.
- Karpinski, D. (2001). So leben sie! Fotoporträts von Familien aus 16 Ländern. Ein Erkundungsprojekt rund um die Welt. Mülheim: Verlag an der Ruhr.
- Katz, P. A. (1973). Stimulus predifferentiation and modification of children's racial attitudes. *Child Development*, 44, 232-237.
- Katz, P. A. (1976). *Towards the elimination of racism*. New York: Pergamon.
- Katz, P. A., Sohn, M., & Zalk, S. (1975). Perceptual concomitants of racial attitudes in urban grade-school children. *Developmental Psychology*, 11, 135-144.
- Katz, P. A., & Zalk, S. R. (1978). Modification of children's racial attitudes. *Developmental Psychology*, 14, 447-461.
- Killen, M., Margie, N. G., & Sinno, S. (2006). Morality in the context of intergroup relationships. In M. Killen & J. Smetana (Eds.), *Handbook of moral development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kinket, B., & Verkuyten, M. (1999). Intergroup evaluations and social context: a multilevel approach. *European Journal of Social Psychology*, 29, 219-237.
- Klasen, H., Woerner, W., Wolke, D., Meyer R., Overmeyer, S., Kaschnitz, W., Rothenberger, A., Goodman, R. (2000). Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. *European Child and Adolescent Psychiatry*, 9, 271-276.

- Klasen, H., Woerner, W., Rothenberger, A., & Goodman, R. (2003). Die deutsche Fassung des Strengths and Difficulties Questionnaire (SDQ-Deu) - Übersicht und Bewertung erster Validierungs- und Normierungsbefunde. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 52(7), 491-502.
- Koslin, S. C., Amarel, M., & Ames, N. (1969). A distance measure of racial attitudes in primary grade children: An exploratory study. *Psychology in the Schools*, 6, 382-385.
- Küchel, J. (2008). *Förderung positiver Intergruppenbeziehungen im Grundschulalter Evaluation einer 'Extended-Contact' Intervention*. Unveröffentlichte Diplomarbeit, Friedrich-Schiller-Universität, Jena.
- Kurdek, L. A. (1977). Structural components and intellectual correlates of cognitive perspective taking in first- through fourth-grade children. *Child Development*, 48, 1503-1511.
- Lessing, E. E., & Clarke, C. (1976). An attempt to reduce ethnic prejudice and to assess its correlates. *Educational Research Quarterly*, 1, 3-16.
- Levy, S. R. (1999). Reducing prejudice: Lessons from social-cognitive factors underlying perceiver differences in prejudice. *Journal of Social Issues*, 55(4), 745-765.
- Lewis, A., & Lewis, V. (1987). The attitudes of young people towards peers with severe learning difficulties. *British Journal of Developmental Psychology*, 5, 287-292.
- Lipscomb, T., McAllister, H. A., & Bregman, N. J. (1985). A developmental inquiry into the effects of multiple models on children's generosity. *Merrill-Palmer Quarterly*, 31(4), 335-344.
- Litcher, J. H., & Johnson, D. W. (1969). Changes in attitudes towards Negroes of white elementary school students after use of multiethnic readers. *Journal of Educational Psychology*, 60, 148-152.
- Litcher, J. H., Johnson, D. W., & Ryan, F. L. (1973). Use of pictures of multiethnic interaction to change attitudes of white elementary school students toward blacks. *Psychological Reports*, 33, 367-372.
- Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1998). Saying no to unwanted thoughts: Self-focus and the regulation of mental life. *Journal of Personality and Social Psychology*, 74(3), 578-589.

- Marcia, J. E. (1980). Identity in adolescence. In J. Adelson (Ed.), *Handbook of adolescent psychology* (159-187). New York: Wiley.
- Marcia, J. E. (1998). Optimal development from an Eriksonian perspective *Encyclopedia of mental health* (Vol. 3, 29-39). New York: Academic Press.
- Marger, M. N. (1991). *Race and ethnic relations*. Belmont, CA: Wadsworth.
- Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology*, 37(3), 431-446.
- McGlothlin, H., Killen, M., & Edmonds, C. (2005). European-American children's intergroup attitudes about peer relationships. *British Journal of Developmental Psychology*, 23, 227-249.
- McGregor, J. (1993). Effectiveness of role playing and antiracist teaching in reducing student prejudice. *Journal of Educational Research*, 86(4), 215-226.
- Messick, D. M., & Mackie, D. M. (1989). Intergroup Relations. *Annual Review of Psychology*, 40, 45-81.
- Milner, D. (1983). *Children and race*. London: Sage.
- Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance and status: An integration. *European Journal of Social Psychology*, 22, 103-122.
- Nachtigall, C., & Wirtz, M. (2002). *Wahrscheinlichkeitsrechnung und Inferenzstatistik*. Weinheim: Juventa-Verlag.
- Nesdale, D. (1999). Social identity and ethnic prejudice in children. In P. Martin & W. Noble (Eds.), *Psychology and society* (92-110). Brisbane: Australian Academic Press.
- Nesdale, D. (2000). Developmental changes in children's ethnic preferences and social cognitions. *Journal of Applied Developmental Psychology*(20), 501-519.
- Nesdale, D. (2001). The development of prejudice in children. In M. Augoustinos & K. Reynolds (Eds.), *Understanding the psychology of prejudice and racism* (57-72). London: Sage.
- Nesdale, D. (2004). Social identity processes and children's ethnic prejudice. In M. Bennett & F. Sani (Eds.), *The development of the social Self*. New York: Psychology Press.

- Nesdale, D. (2008). Social identity development and children's ethnic attitudes in Australia. In S. M. Quintana & C. Mckown (Eds.), *Handbook of race, racism and the developing child* (313-338). Hoboken, New Jersey: John Wiley & Sons.
- Nesdale, D., Durkin, K., Maas, A., & Griffiths, J. (2005). Group norms, threat, and children's racial prejudice. *Child Development*, 76(3), 652-663.
- Nesdale, D., & Flessner, D. (2001). Social identity and the development of children's group attitudes. *Child Development*, 72, 506-517.
- Nesdale, D., Maass, A., Griffiths, J., & Durkin, K. (2003). Effects of ingroup and outgroup ethnicity on children's attitudes towards members of the ingroup and outgroup *British Journal of Developmental Psychology*, 21, 177-192.
- Nisbett, R. E., Krantz, D.H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90, 339-363.
- Nitsche, S. (2008). *Die Entwicklung von Vorurteilen: Eine längsschnittliche und multimethodale Betrachtung der Veränderungen innerhalb des dritten Schuljahres*. Unveröffentlichte Diplomarbeit, Friedrich-Schiller-Universität, Jena.
- Offord, D. R. (2000). Prevention of antisocial personality disorder. In J. L. Rapoport (Ed.), *Childhood onset of 'adult' psychopathology: Clinical and research advances* (379-398). Arlington, VA: US: American Psychiatric Publishing, Inc.
- Oppenheimer, L. (2006). The development of enemy images: A theoretical contribution. *Journal of Peace and Conflict*, 12, 269-292.
- Osborne, J. W. (1997). Race and academic disidentification. *Journal of Educational Psychology*, 4, 728-735.
- Oskamp, S. (1991). *Attitudes and opinions* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Oskamp, S. (2000). Multiple paths to reducing prejudice and discrimination. In S. Oskamp (Ed.), *Reducing prejudice and discrimination*. Mahwah: Lawrence Erlbaum Associates.
- Perkins, D. M., & Mebert, C. J. (2005). Efficacy of multicultural education for preschool children: A domain-specific approach. *Journal of Cross-Cultural Psychology*, 36(4), 497-512.
- Peters, W. (1970). *The eye of the storm*. USA.

- Pettigrew, T. F. & Meertens, R. W. (1995). Subtle and blatant prejudice in Western Europe. *European Journal of Social Psychology*, 25(1), 57-75.
- Pettigrew, T. F., & Tropp, L. R. (2006). A Meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751-783.
- Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology*, 38, 922-934.
- Pfeifer, J. H., Spears-Brown, C., & Juvonen, J. (1999). Teaching tolerance in schools: Lessons learned since Brown v. Board of Education about the development and reduction of children's prejudice. *in Druck*.
- Piaget, J. (1932). *The moral judgement of the child*. London: Kegan Paul.
- Piaget, J., & Weil, A.-M. (1951). The development in children of the idea of the homeland and of relations with other countries. *International Social Science Bulletin*, 3, 561-578.
- Quaquebeke, N. v., Henrich, D. C., & Eckloff, T. (2007). "It's not tolerance I'm asking for, it's respect!" A conceptual framework to differentiate between tolerance, acceptance and (two types of) respect. *Gruppendynamik und Organisationsberatung*, 38(2), 185-200.
- Quattrone, G. A., & Jones, E. E. (1980). The perception of variability within in-groups and out-groups: Implications for the law of small numbers. *Journal of Personality and Social Psychology*, 38, 141-152.
- Raabe, T., & Beelmann, A. (in press). Development of ethnic, racial, and national prejudice in childhood and adolescence: A multinational meta-analysis of age differences. *Child Development*.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-Analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59-92.
- Rubin, K. H. (1978). Role taking in childhood: Some methodological considerations. *Child Development*, 49, 428-433.
- Rutland, A. (1999). The development of national prejudice, in-group favouritism and self-stereotypes in British children. *British Journal of Social Psychology*, 38, 55- 70.

- Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development, 76*(2), 451-466.
- Sameroff, A. J., & Fiese, B. H. (2000). Transactional regulation: The developmental ecology of early intervention. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (135-159). Cambridge: Cambridge University Press.
- Sani, F., Bennett, M., Agostini, L., Malucchi, L., & Ferguson, N. (2000). Children's conceptions of characteristic features of category members. *Journal of Social Psychology, 140*, 227-239.
- Schopler, J., Insko, C. A., Drigotas, S., & Graetz, K. A. (1993). Individual-group discontinuity: Further evidence for mediation by fear and greed. *Personality and Social Psychology Bulletin, 19*(4), 419-431.
- Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1961). *Intergroup conflict and cooperation: The robbers cave experiment*. Norman, OK: University Book Exchange.
- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York, NY, US: Cambridge University Press.
- Signorella, M. L., & Liben, L. S. (1985). Assessing children's gender stereotyped attitudes. *Psychological Documents, 15*, 7.
- Sinclair, S., Dunn, E., & Lowery, B. S. (2005). The relationship between parental racial attitudes and children's implicit prejudice *Journal of Experimental Social Psychology 41*, 283- 289.
- Singh, J. M., & Yancey, A. V. (1974). Racial attitudes in white first-grade children. *Journal of Educational Research, 67*(8), 370-372.
- Stephan, C. W., Renfro, L., & Stephan, W. G. (2004). The evaluation of multicultural education programs: Techniques and a meta-analysis. In W. G. Stephan & W. P. Vogt (Eds.), *Education programs for improving intergroup relations*. New York: Teachers College Press.
- Stephan, W. G., & Finlay, K. (1999). The role of empathy in improving intergroup relations. *Journal of Social Issues, 55*(4), 729-743.
- Stephan, W. G., & Stephan C.W. (1985). Intergroup Anxiety. *Journal of Social Issues, 41*, 157-175.

- Stephan, W. G., & Stephan, C. W. (2001). *Improving intergroup relations*. Thousand Oaks, CA, US: Sage Publications.
- Steudel, A. (2010). *Perspektivenübernahme als Prädiktor von Vorurteilen und Toleranz? Eine längsschnittliche Analyse im Grundschulalter*. Unveröffentlichte Diplomarbeit, Friedrich-Schiller-Universität, Jena.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations*. Monterey, CA: Brooks/Cole.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations*. Chicago: Nelson Hall.
- Tauran, R. H. (1967). *The influences of reading on the attitudes of third graders toward Eskimos*. Unpublished doctoral dissertation, University of Maryland, College Park.
- Tausch, N., Hewstone, M., Kenworthy, J., Cairns, E., & Christ, O. (2007). Cross-community contact, perceived status differences, and intergroup attitudes in Northern Ireland: The mediating roles of individual-level versus group-level threats and the moderating role of social identification. *Political Psychology*, 28, 53-68.
- Teichman, Y. (2001). The development of Israeli children's images of Jews and Arabs and their expression in human figure drawings. *Developmental Psychology*, 37, 749-761.
- Teichman, Y., & Bar-Tal, D. (2008). Acquisition and development of a shared psychological intergroup repertoire in a context of an intractable conflict. In S. M. Quintana & C. McKown (Eds.), *Handbook of race, racism and the developing child* (452-482). Hoboken, New Jersey: John Wiley & Sons.
- Thüringer Innenministerium (2009). *Verfassungsschutzbericht Freistaat Thüringen 2008*.
- Trager, H. G., & Yarrow, M. R. (1952). *They learn what they live*. New York: Harper.
- Triandis, H. C. (1972). *The analysis of subjective culture*. New York: Wiley.

- Tropp, L. R., & Prenovost, M.A. (2008). The role of intergroup contact in predicting children's interethnic attitudes: Evidence from meta-analytic and field studies. In S. R. Levy, & M. Killen (Ed.), *Intergroup attitudes and relations in childhood through adulthood*. Oxford: University Press.
- Turner, J. C., Hogg, M.A., Oakes, P.J., Reicher, S.D., & Wetherell, M.S. (1987). *Rediscovering the social group: A self-categorization theory*. New York: Basil Blackwell.
- Turner, R. N., Hewstone, M., & Voci, A. (2007). Reducing explicit and implicit outgroup prejudice via direct and extended contact: The mediating role of self-disclosure and intergroup anxiety. *Journal of Personality and Social Psychology*, 93(3), 369-388.
- Valk, A., & Karu, K. (2001). Ethnic attitudes in relation to ethnic pride and ethnic differentiation. *The Journal of Social Psychology*, 141, 583-601.
- Verkuyten, M. (2001). National identification and intergroup evaluations in Dutch children. *British Journal of Developmental Psychology*, 19, 559-571.
- Verkuyten, M. (2002). Ethnic attitudes among minority and majority children: The role of ethnic identification, peer group victimization and parents. *Social Development*, 11, 558-570.
- Vygotsky, L. S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Weiner, M. J., & Wright, F. E. (1973). Effects of undergoing arbitrary discrimination upon subsequent attitudes toward a minority group. *Journal of Applied Social Psychology*, 3, 94-102.
- Westphal, R. C. (1977). *The effects of primary grade level interethnic curriculum on racial prejudice*. California: Rand & Research Associates.
- Whetstone, L. M. (1992). *An evaluation of prejudice reduction program for children*. Unpublished dissertation, Claremont Graduate School, CA.
- Williams, J. E., Best, D. L., Boswell, D. A., Mattson, L. A., & Graves, D. J. (1975). Preschool racial attitude measure II. *Educational and Psychological Measurement*, 35, 3-18.
- Wittenbrink, B., & Henly, J. R. (1996). Creating social reality: Informational social influence and the content of stereotypic beliefs. *Personality and Social Psychology Bulletin*, 22(6), 598-610.

- Wright, S. C., Aron, A., McLaughlin-Volpe, T., Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology*, 73(1), 73-90.
- Yawkey, T. D. (1973). Attitudes toward black Americans held by rural and urban white early childhood subjects based upon multiethnic social studies materials. *The Journal of Negro Education*, 42, 164-169.
- Zalk, S. R., & Katz, P. A. (1976). The Katz-Zalk projective prejudice test: Measure of racial attitudes in children. *Catalog of Selected Documents in Psychology*, 6, 7-38.
- Ziegler, S. (1981). The effectiveness of cooperative learning teams for increasing cross-ethnic friendship: Additional evidence. *Human Organization*, 40, 264-268.

Lebenslauf**Michael Saur****Diplom-Psychologe**

Geboren am 17.10.1977 in Datteln

Ledig

Römisch-katholisch

Ausbildung

1985-1988	Grundschule: Lohschule Datteln
1988-1997	Gymnasium: Comenius Gymnasium Datteln
05.06.1997	Abschluss: Abitur
1998-2005	Friedrich-Schiller-Universität Jena Hochschulstudium der Psychologie mit Schwerpunkt klinischer und Arbeits-, Betriebs- und Organisationspsychologie
WS 2000/2001	Freie Universität Amsterdam Auslandssemester
01.07.2005	Abschluss: Diplom Psychologe
2005-2010	DFG-Forscherguppe Diskriminierung und Toleranz in Intergruppenbeziehungen an der FSU Jena Wissenschaftlicher Mitarbeiter im Projekt Förderung von Intergruppenbeziehungen bei Kindern

Ort, Datum

Unterschrift

Ehrenwörtliche Erklärung

Ich erkläre hiermit, dass mir die geltende Promotionsordnung der Fakultät für Sozial- und Verhaltenswissenschaften der Friedrich-Schiller-Universität Jena bekannt ist.

Ferner erkläre ich, dass ich die vorliegende Arbeit selbst und ohne unzulässige Hilfe Dritter angefertigt habe. Alle von mir benutzten Hilfsmittel, persönliche Mitteilungen und Quellen sind in der Arbeit angegeben.

Bei der formalen Korrektur des Manuskriptes (Rechtschreibung und Grammatik) haben mich folgende Personen unentgeltlich unterstützt: Gerda Saur, Reiner Saur, Michael Hock, Kevin Faber, Martha Pobering, Kerstin Rische, Elmar Egner und Anna Steudel.

Insbesondere habe ich für diese Arbeit nicht die Hilfe eines Promotionsberaters in Anspruch genommen, und Dritte haben weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Weder früher, noch gegenwärtig habe ich an einer anderen Hochschule die Dissertation als Prüfungsarbeit für eine staatliche oder wissenschaftliche Prüfung eingereicht.

Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Ort, Datum

Unterschrift